

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 10, 2002, 11:09:23 ; Search time 28.49 Seconds

(without alignments)  
2143.461 Million cell updates/sec

Title: US-09-905-075-2

Perfect score: 2005

Sequence: 1 MRLPRAAGLGLPLLLLP.....AEAEATGESPTQLPSREDL 353

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : SPREMBL\_19:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_plant:  
10: sp\_protist:  
11: sp\_virus:  
12: sp\_vertebrate:  
13: sp\_unclassified:  
14: sp\_virus:  
15: sp\_bacteriophage:  
16: sp\_archaeo:  
17: sp\_archaeo:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4787	89.1	321	4	Q9BU47
2	1535.5	76.5	350	11	Q9CYA0
3	1481.5	73.9	348	11	Q60438
4	969	48.3	422	4	Q9BHD1
5	968.5	48.3	420	11	Q91XD7
6	959.5	47.9	417	4	Q9Y409
7	670	33.4	374	5	Q9VPD0
8	599.5	29.9	165	13	Q9DEE9
9	540	26.9	356	5	Q19267
10	286.5	14.3	704	13	Q73774
11	279.5	13.9	937	5	Q9BLJ1
12	272	13.6	1587	4	Q00508
13	271.5	13.5	1511	4	Q75412
14	270.5	13.5	2189	5	Q9BI05
15	270.5	13.5	3857	11	Q88840
16	268	13.4	1821	4	Q14767

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	321 AA.
Q9BU47	Q9BU47			
AC	Q9BU47			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	HOMO SAPIENS 35.0 KDA PROTEIN.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
NCBI	Taxid=9606;			
RP	SEQUENCE FROM N. A.			
RC	TISSUE=LUNG CARCINOMA;			
RA	Straussberg R.;			
RL	Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC002894; AA02894.1; -			
DR	HSP; P35555; IEMN.			

DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR SMART: SM00181; EGF; 3.  
 DR SMART: SM00179; EGF\_CA; 2.  
 DR SMART: SM00001; EGF\_Like; 1.  
 DR SMART: SM00261; FU; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01187; EGF\_2; 2.  
 DR PROSITE: PS01187; EGF\_CA; 2.  
 DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 DR Calcium-binding: EGF-like domain; Glycoprotein; Hydroxylation;  
 KW Hypothetical protein; Repeat.  
 SO SEQUENCE 321 AA; 34958 MW; A74360A1D817F23D CRC64;

Query Match 89.1%; Score 1787; DB 4; Length 321;  
 Best Local Similarity 90.9%; Pred. No. 9,1e-164;  
 Matches 321; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MRLPRRAALGLPLLLPPAPAAKPPCHRCGLVDKFNQGMVDTAKNFGGNTAW 60  
 DB 1 MRLPRRAALGLPLLLPPAPAAKPPCHRCGLVDKFNQGMVDTAKNFGGNTAW 60  
 QY 61 EEKTLKSESEIRLEILGLCESSDFECNQMLEAOEHLAMWTLQKSEYDLPFEMFC 120  
 DB 61 EEKTLKSESEIRLEILGLCESSDFECNQMLEAOEHLAMWTLQKSEYDLPFEMFC 120  
 QY 121 VKTLKCCSPGTGPPCLACQGGSORPCSGNGHCSGDSHQDGGSCRHMGYGGPLCTDC 180  
 DB 121 VKTLKCCSPGTGPPCLACQGGSORPCSGNGHCSGDSHQDGGSCRHMGYGGPLCTDC 180  
 QY 181 MDGFSSLRNETHSICITACDESKCTSGLTNRCCGCEVGVWVLDGACVYVDECAEAPP 240  
 DB 181 MDGFSSLRNETHSICITACDESKCTSGLTNRCCGCEVGVWVLDGACVYVDECAEAPP 240  
 QY 241 CSAAOFCKNANGSYTCEDCSSCVGCTGEGPGNCKECISGYAREHGCADVDDECSLAEXT 300  
 DB 241 CSAAOFCKNANGSYTCEDCSSCVGCTGEGPGNCKECISGYAREHGCADVDDECSLAEXT 300  
 QY 301 CVRKNENCVTPGSIYCVCPDGFETEDACVPAAEAATGESPTQLPSREDL 353  
 DB 269 CVRKNENCVTPGSIYCVCPDGFETEDACVPAAEAATGESPTQLPSREDL 321

RESULT 2  
 Q9CYAO PRELIMINARY; PRT; 350 AA.  
 AC 09CYAO;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE 5730592L2LRK PROTEIN.  
 GN 5730592L2LRK.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Balzarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gattung S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazarrelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK017880; BAB30986.1; -.  
 DR HSP; P35555; IEMN.  
 DR MGI:1923987; 5730592L2LRK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR002174; Furin-like.  
 DR SMART: SM00181; EGF; 4.  
 DR SMART: SM00179; EGF\_CA; 3.  
 DR SMART: SM00001; EGF\_Like; 1.  
 DR SMART: SM00261; FU; 3.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 KW EGF-like domain; Glycoprotein; Hydroxylation;  
 SO SEQUENCE 350 AA; 38219 MW; 781D7389B1944231 CRC64;

Query Match 76.5%; Score 1533.5; DB 11; Length 350;  
 Best Local Similarity 75.6%; Pred. No. 2.4e-139;  
 Matches 267; Conservative 30; Mismatches 53; Indels 3; Gaps 2;

QY 1 MRLPRRAALGLPLLLPPAPAAKPPCHRCGLVDKFNQGMVDTAKNFGGNTAW 60  
 DB 1 MRLPRRAALGLPLLLPPAPAAKPPCHRCGLVDKFNQGMVDTAKNFGGNTAW 58  
 QY 61 EEKTLKSESEIRLEILGLCESSDFECNQMLEAOEHLAMWTLQKSEYDLPFEMFC 120  
 DB 59 EEKTLKSESEIRLEILGLCESSDFECNQMLEAOEHLAMWTLQKSEYDLPFEMFC 118  
 QY 121 VKTLKCCSPGTGPPCLACQGGSORPCSGNGHCSGDSHQDGGSCRHMGYGGPLCTDC 180  
 DB 119 VKTLKCCSPGTGPPCLACQGGSORPCSGNGHCSGDSHQDGGSCRHMGYGGPLCTDC 178  
 QY 181 MDGFSSLRNETHSICITACDESKCTSGLTNRCCGCEVGVWVLDGACVYVDECAEAPP 240  
 DB 179 MDGFSSLRNETHSICITACDESKCTSGLTNRCCGCEVGVWVLDGACVYVDECAEAPP 238  
 QY 241 CSAAOFCKNANGSYTCEDCSSCVGCTGEGPGNCKECISGYAREHGCADVDDECSLAEXT 300  
 DB 239 CSAAOFCKNANGSYTCEDCSSCVGCTGEGPGNCKECISGYAREHGCADVDDECSLAEXT 298  
 QY 301 CVRKNENCVTPGSIYCVCPDGFETEDACVPAAEAATGESPTQLPSREDL 353  
 DB 299 CVRKNENCVTPGSIYCVCPDGFETEDACVPAAEAATGESPTQLPSREDL 350

RESULT 3  
 Q60438 PRELIMINARY; PRT; 348 AA.  
 AC Q60438;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HT PROTEIN  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OX Cricetus.  
 OX NCBI\_TaxId=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Chen H., Okubo, T., Ling V., Zhang W.;



```

OY 286 GOCADVDECSLAERTKVRNENCYNTPGSYVCVCPDGFETEDAC---VPAE---ABA 338
DB 301 SCSIDVDECEV--VCPGENKCENTEGGYRCVACAGYHQEDGICVKEQVPESAGFFAM 358
OY 339 TEGE 342
DB 359 TEDE 362

RESULT 6
OY409 PRELIMINARY: PRT: 417 AA.
AC 09Y409:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HYPOTHEICAL. 44.9 KDA PROTEIN.
GN DKFZP566D213.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA MEDLINE=21154917; PubMed=11230166;
RA Wieman S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansgore W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Meves H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wamutts R., Korn B., Klein M., Poustka A.;
RA Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.;
RL Genome Res. 11:422-435(2001).
DR EMBL: AL050275; CAB4376.1; -.
DR HSSP: P35555; 1EMN.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002049; Laminin_EGF.
DR SMART: SM00181; EGF_1.
DR SMART: SM00001; EGF_Like: 1.
DR SMART: SM00261; FU; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 2.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hypothetical protein;
KW Repeat.
KM SEQUENCE 417 AA: 44934 MW; C2E38C3714E5C4B8 CRC64;
SO

Query Match 47.9%; Score 959.5; DB 4; Length 417;
Best Local Similarity 47.9%; Pred. No. 4.7e-84;
Matches 169; Conservative 47; Mismatches 118; Indels 19; Gaps 5;

```

```

OY 237 EPPCSAOFCKNANGSYTCERDSSVCCTGEGPNCCKECISGVAREHGOCADVDECSL 296
DB 252 EGANCGADFCVNTBESSYICRCACACACAGMGAPRCCKCSGQVQVQSKCLADVDECE 310
OY 297 AEKTCVRKNENCYNTPGSYVCVCPDGFETEDACV---PAEA---EATGE 342
DB 311 -TEVCPENKOCENTEGGYRCICAECKOMEKICVKEQIPESAGPSEMTED 362

RESULT 7
OY409 PRELIMINARY: PRT: 374 AA.
AC 09Y409:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CG11377 PROTEIN.
GN CG11377.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Gerhart M., Gelfand S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003590; AAF51560.1; -.
DR FlyBase: FBgn0031217; CG11377.
DR InterPro: IPR000561; EGF-Like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002049; Laminin_EGF.
DR SMART: SM00181; EGF_1.

```



DR SMART; SM00001; EGF\_like; 1.  
 DR SMART; SM00261; FU; 2.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.  
 SQ SEQUENCE 374 AA; 41161 MW; 6B093AFC28331B0D CRC64;

Query Match 33.4%; Score 670; DB 5; Length 374;  
 Best Local Similarity 40.5%; Pred. No. 2,8e-56;  
 Matches 122; Conservative 48; Mismatches 109; Indels 22; Gaps 8;

QY 28 PTCHRCRGVLDKFNQGMVDTAKKNGGNTAMEKTLSESEITLLELDELCSSD 87  
 DB 28 PPRCRACTQLVSSFRAL-ERTKRGHAGDPTAMEEKLRSKNEVRLVEOEKLCSEGE 86  
 QY 88 F---ECNOMLEAOEHLAEAMWLQKSEYDLEFEMFCVTKLVKCCSPGTYGPDCLACGG 143  
 DB 87 VINKDHCHNLANEHEALLEDMFIHKOTESPDLQSLCIDQITVCCPLNTYGPDLCTC--- 143  
 QY 144 SQPCSGNGHSGDGSFGDSCRCCHMGYOGPLCTDCMDGYFSSLRNETHSICTACDESC 203  
 DB 144 --TECNNGKCKGDKTRKNGKCKCDPGYAGPNCNECGPEHYESFRDEKKLLCTQCHAC 201  
 QY 204 KT--CSGLTRNDGCECEVGVLD-EGACVDVDECAE--PPCSAAPCKNANSYTCCE 258  
 DB 202 GEGGCTGEGGKSRCKCKGKMSMEACVDINECLEQQRPPCKPQGFVNNESFSFCL 261  
 QY 259 CDSGVCCTGEGGKCKECSISGAREHGQCADVDECSLAERTCYRKNENCVNTPGSYVCV 318  
 DB 262 CDRSCDCCDDGDPKMKCKADGYELKRGKCHDIS---AEO---RSNVSTRLALYTFGM 314  
 QY 319 C 319  
 DB 315 C 315

RESULT 8  
 Q9DFE9 PRELIMINARY; PRT; 165 AA.  
 AC Q9DFE9;  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 17, Last sequence update)  
 DE HYPOHETICAL 18.7 KDA PROTEIN (FRAGMENT).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 NC NCBITaxID=8022;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bayne C.J., Gerwick L., Fujiki K., Nakao M., Yano T.;  
 RT "Immune-relevant (including acute phase) genes identified in the  
 RT livers of rainbow trout, Oncorhynchus mykiss, by means of suppression  
 RT subtractive hybridization."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF281341; AAG30016.1;  
 DR InterPro; IPR000561; EGF-like.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00001; EGF\_like; 1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 KW Hypothetical protein.  
 FT NON\_TER 1 165  
 SQ SEQUENCE 165 AA; 18682 MW; A93BE87C38CF4502 CRC64;

Query Match 29.9%; Score 599.5; DB 13; Length 165;  
 Best Local Similarity 57.0%; Pred. No. 6.7e-50;  
 Matches 94; Conservative 28; Mismatches 42; Indels 1; Gaps 1;

QY 82 LCSSDFECNOMLEAOEHLAEAMWLQKSEYDLEFEMFCVTKLVKCCSPGTYGPDCLACQ 141  
 DB 1 LCSSDFECNOMLEAOEHLAEAMWLQKSEYDLEFEMFCVTKLVKCCSPGTYGPDCLACQ 60  
 QY 142 GGSORPCSGNGHSGDGSFGDSCRCCHMGYOGPLCTDCMDGYFSSLRNETHSICTACDE 201  
 DB 61 GGSERPCGHNGVCGDTRGNGNRCNDHGYKGFCLDCMDGYFSSLRNETHSICTACDE 120  
 QY 202 SCTKCSGLTRNDGCECEVGV-VIDEGACVDVDECAEPPCSAAQ 245  
 DB 121 SCTKCTGTMTNEDCKCKTGMGEDDERLTCLVDINECLNDPPLCKEDQ 165

RESULT 9  
 Q19267 PRELIMINARY; PRT; 356 AA.  
 AC Q19267;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE F09E8.2 PROTEIN.  
 GN F09E8.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NC NCBITaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Percy C.M.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99065613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z73896; CA98055.1;  
 DR InterPro; IPR000152; ASX\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF\_CA.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR000004; Sapp.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00001; EGF\_like; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SQ SEQUENCE 356 AA; 39790 MW; EFF82577DE334F57 CRC64;

Query Match 26.9%; Score 540; DB 5; Length 356;  
 Best Local Similarity 36.1%; Pred. No. 8.5e-44;  
 Matches 108; Conservative 43; Mismatches 110; Indels 38; Gaps 6;  
 QY 11 LPLPLLLPPAPFAAKPTPCRHRCGLVDKFNQGMVDTAKKNGGNTAMEKTLSEKYES 70  
 DB 7 LLAVLIGATSGKVTINNEKRCNTCNFLVSTFDEGLAKTARHNFAGDPTAMEEKLRSK 66  
 QY 71 SEIRLLETLEGLCESS-----DFECNOMLEAOEHLAEAMWLQKSEYDLEF 116  
 DB 67 SETRLIEVLEGVCKKSLPPNNDNFMGIAIEFKCSTOLEKHEETIEFY--YNQOHNMYS 124  
 QY 117 EMPFCVTKLVKCCSPGTYGPDCLACQGSORP--CSGNGHSGDGSFGDSCRCCHMGYOG 174  
 DB 125 NWLCEVQDLKLCPPGHRKCKNEQCPGLSEKADVCFGKSGCHDGSREGSKCKETGTG 184  
 QY 175 PLCTDCMDGYFSSLRNETHSICTACDESC-KTCSGLTRNDGCECEVGVLDVDEGACVDVDE 233

```

Db 185 NLCRYDIEFEESRTVOGVYCKKHGBCLGVSSSESSKCKCKNGWKLTEEGCADVNE 244
OY 234 CAAEPPCSAAQPCFKNANGSYTCEDSSCGCTGEGPNCCKICISGAYAREHQCADVD 292
Db 245 CONESACTKHELCVMTVGSFKCE-----CKE---GYKKDDQONQOFD 284

RESULT 10
ID 073774 PRELIMINARY; PRT; 704 AA.
AC 073774:
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE FIBULIN-1, ISOFORM D.
GN FBLN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120531; PubMed=9923656;
RA Barth J.L., Agraves K.M., Roark E.F., Little C.D., Agraves W.S.;
RT "Identification of chicken and C. elegans fibulin-1 homologs and
RT characterization of the C. elegans fibulin-1 gene.";
RL Matrix Biol. 17:635-646(1998).
DR EMBL: AF051399; AAC05387.1; -.
DR HSSP: P00742; IHCG.
DR InterPro: IPR000020; Anaphylatoxin.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF01821; ANATO; 2.
DR Pfam: PF00008; EGF_6.
DR SMART: SM00104; ANATO; 3.
DR SMART: SM00179; EGF_CA; 8.
DR SMART: SM00001; EGF_Like; 1.
DR PROSITE: PS01177; ANAPHYLATOXIN_1; UNKNOWN_1.
DR PROSITE: PS01178; ANAPHYLATOXIN_2; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; 5.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS01187; EGF_CA; 8.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 704 AA; 78137 MW; D47D5A30D5E42932 CRC64;
SQ

Query Match 14.3%; Score 286.5; DB 13; Length 704;
Best Local Similarity 25.7%; Pred. No. 4,4e-19;
Matches 127; Conservative 37; Mismatches 120; Indels 211; Gaps 29;
OY 6 RAALGLPLLLLPAPPAEAKKPTPCRCGLVKNQGVADPAKKNGFGGNTAMEKTL 65
Db 5 RGRAPRLLLLLALLPALRGDLSMEEC---CDK---GVENANKN----- 44
OY 66 SKYESSEIRLLELEGLICSSDFECNQMLEAQ-----BEHLEAWMLQIKSEYPDL---- 115
Db 45 -----RICTSLPLISER--ECS-MTQVQCCKSLKEHYHSCDGLIEFASVHEEDSHN 93
OY 116 -----FEMF-----CVTLKV-----CSPGTYGPDCL- 138
Db 94 GENSTCEAEYFKKCCYCCLLGKTAOVQOSCEFNLEKIGYQCGIVFRACVKGQEGTDVSI 153
OY 139 -----ACOGGSGRPGSGNGHSGDGSROGDSRCRCRMGY 172
Db 154 SDDAPKKEQVEISKEDLDQEDPLHNGCRGG--PCS--QQCNDTSSY---VCSFVGI 206
OY 173 Q-----GPLCTD---CMQGYFSSLRNETHS--ICTAC-----DESCKCSGLTN- 211
Db 207 QIQPDPVNCEDINECITG-----THSGIGQTCVNTLGSFRQGRDRDSCGIGYELTLD 258
OY 212 ---RDCGECEV-----GWLIDE--GACVDVDECAEPP 239

```

```

Db 259 SRCKDIDECETGTHNCPPDFICONTGSPRCRKLQCMNGFIODALGNCIDINECLSTNN 318
OY 240 PCSAAQFCCKNANGSYTC-----EECDSSCVCTGE-----G 270
Db 319 PCPAGQICNTDGSYTCORISPCGRCGYHLNEDGTACVDVBEGSSDOPC-GEHVCING 377
OY 271 PGNCK-ECISGARE--HGOCADVDEC-SLAERTCYRKNECNYTPGYSVYVCPPDGEET 326
Db 378 PGNYRCECKSGSYFVDYISRCIDINECRYPGRCAHK---CENTPGSYVCTCTMFKLS 434
OY 327 ED--ACVPAEAEAT 339
Db 435 SDGRSCEDLINECSS 449

RESULT 11
ID 09BLJ1 PRELIMINARY; PRT; 937 AA.
AC 09BLJ1:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CI-METAL.
GN CI-METAL.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;
OC Clonidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21347414; PubMed=11455433;
RA Nakayama A., Satou Y., Satou N.;
RT "Isolation and characterization of genes that are expressed during
RT Ciona intestinalis metamorphosis.";
RL Dev. Genes Evol. 211:184-189(2001).
DR EMBL: AB041857; BAB40596.1; -.
DR HSSP: P00742; IHCG.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR00742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001212; Somatomedin_B.
DR InterPro: IPR001491; Thrombomodulin.
DR Pfam: PF00008; EGF_5.
DR PRINTS: PR00907; THROMBOMODULN.
DR SMART: SM00181; EGF_19.
DR SMART: SM00179; EGF_CA; 18.
DR PROSITE: PS00010; ASX_HYDROXYL; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 19.
DR PROSITE: PS01187; EGF_CA; 12.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 937 AA; 101043 MW; BC67B30C8E391D07 CRC64;
SQ

Query Match 13.9%; Score 279.5; DB 5; Length 937;
Best Local Similarity 29.3%; Pred. No. 2,9e-18;
Matches 80; Conservative 24; Mismatches 96; Indels 73; Gaps 14;
OY 121 VTKLVCC-SPTGY-----GPDLACGGGSGRPGSGNGHSGDGSROGDSRCRC 168
Db 461 VMTNKRRCINTPGSFRVCNNGYRAQGSRCVDINECRSSPCGNNAKRI---NTPGSFTCRK 517
OY 169 HMGY--QGFLCTDCMDGYSSLRNETHSICTACDECKCSGLTNRDCECEYGVWLIDG 226
Db 518 NTGYTGNGLICRDI-----NECEAHPNPCGENALCINGISYRC-ICARGF--SGP 565
OY 227 ACVDVDECAEPPCSAAQFCCKNANGSYTCEDSSCGCTGEG- 270
Db 566 LCTDFNECAAIRPCSPNADCTNTGSPFC-QCKP---GYTGGLVCRDINECSRPNACP 621

```

QY	271	-----PG--NCKECJSGVAREHGOCADVDECSLAERTCYVRKKNENINPGSGVYVCP	320
Db	622	RNQCINTPGGFNC-VCAITGRVYRGNCVNDINECRASRRPC-DLNASCQNPGPSFTCTCN	679
QY	321	DGFE-----ETEDACVPPEAEATEEG	341
Db	680	TCYTGNGLTCAADINECNNPRACHPQATCANTPG	712
RESULT 12			
ID	000508	PRELIMINARY:	PRT: 1587 AA.
AC	000508:		
DT	01-JUL-1997	(TREMBLrel. 04, Created)	
DT	01-JUL-1997	(TREMBLrel. 04, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	LATENT TGF-BETA BINDING PROTEIN-4.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97415399; PubMed=9271198;		
RA	Giltay R., Kostka G., Timp R.;		
RT	"Sequence and expression of a novel member (LBP-4) of the family of		
RL	latent transforming growth factor-beta binding proteins.";		
RL	FEBS Lett. 411:164-168(1997).		
DR	EMBL: Y13622; CAI73944.1; "		
DR	HSSP: P35555; IEMN		
DR	InterPro: IPR000152; Asx_hydroxyl.		
DR	InterPro: IPR000561; EGF-like.		
DR	InterPro: IPR001881; EGF_Ca.		
DR	InterPro: IPR003018; GAF.		
DR	InterPro: IPR002212; TB.		
DR	Pfam: PF00008; EGF_17.		
DR	Pfam: PF00683; TB; 4.		
DR	SMART: SM00179; EGF_CA; 15.		
DR	SMART: SM00001; EGF-like; 5.		
DR	SMART: SM00065; GAF; 1.		
DR	PROSITE: PS00010; ASX_HYDROXYL; 14.		
DR	PROSITE: PS00027; EGF_1; UNKNOWN_2.		
DR	PROSITE: PS01186; EGF_2; 12.		
DR	PROSITE: PS01187; EGF_CA; 17.		
KW	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.		
SC	SEQUENCE 1587 AA; 169620 MW; 57AB3295FPA0AD46 CRC64;		
QY	Query Match 13.6%; Score 272; DB 4; Length 1587;		
QY	Best Local Similarity 24.0%; Pred. No. 2.8e-17;		
QY	Matches 113; Conservative 27; Mismatches 122; Indels 208; Gaps 27;		
QY	19	PPAP-----EAAKKRTPC--HRCRLVLDKEN-----QGWDTAKKNFGGNTAMEKTLISK	67
Db	709	PCAPQADYDECKSRSPPCYGRCENTEGSFQCVCPMGFOP-----NTAGSE-----	754
QY	68	YSESEIRLEILEIEGLCESSDFECNOMLEAQEENLEAMWLQKSEYPDLEFMCVTKLVC	127
Db	755	-----CEVD-EC-----EHLACPOGE-----CVN-----	774
QY	128	CSPTGYRPGDCLACGG-----SQRPCSGNGHCSGDSHSGDGSRCRHMG	171
Db	775	-SPGSF--QCRTCPGSHLHGRGCTDVDECCSGAPPCGPHGCT--NTEGSGFSCAPG	828
QY	172	YGGPL-----CTD-----CMDGYF-----SLRRETHSICTA-----CD	200
Db	829	YRAPSGRGPPCADVNECLEGGDFCPFHGECLNTDGSFACTACAPGRPRGASCILDVDECS	888
QY	201	ESCKTCSGL--TNRD-----CG-----	215
Db	889	EEIDLQSGSIGCTNTDSFEDICPPHRRAGDLASCLDVDECRGRGALGCSQRCENSPGSI	948
QY	216	----BCEVGM-VLDGRACVDVDECAAEPPPCGSAOFCCKNANGSYTC-EECDSSVCCTGE	269

```

Db      949 KCVNRDDEPGHAGPSTCTDDECOEYGPETICGAGQRCENTPGSYRTPACPDGEYOTPTP 1007
QY      270 GPGNCKE-----CISGY--AREHGOCADVDECSLAEKTC 301
Db      1008 --GGCQDVDECCNRBFCCGAAVCONPGSFQCLCDQGYGADGRICVADVNEETLQAGVC 1065
QY      302 VAKNENCYTPGSYVCVCPDGFETE--DACVPPAEATRGESPTQLP 348
Db      1066 --GAALCEWVESSFLVCPCNSPEEFDPMGRGVPP--RTSVGMSPGSGP 1110

RESULT 13

ID 075412 PRELIMINARY: PRT, 1511 AA.
AC 075412;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE LATENT TRANSFORMING GROWTH FACTOR-BETA BINDING PROTEIN 45.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA MEDLINE=98325059; PubMed=9660815;
RX Saharinen J., Taipale J., Monti O., Keski-Oja J.;
RT "Identification and characterization of a new latent transforming
RL growth factor-binding protein, LTRP-4."
DR EMBL: AF051344; AAC39879.1; -.
DR HSSP: P35555; IEMN.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR Pfam: PF00008; EGF_17.
DR Pfam: PF00683; TB_4.
DR SMART: SM00179; EGF_CA_15.
DR SMART: SM00001; EGF_Like_5.
DR PROSITE: PS00010; ASX_HYDROXYL_14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2_12.
DR PROSITE: PS01187; EGF_CA_17.
KM Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 1511 AA; 161157 MW; C61AB75B265958D CRC64;

Query Match 13.5%; Score 271.5; Db 4; Length 1511;
Best Local Similarity 23.6%; Pred. No. 2,9e-17;
Matches 110; Conservative 29; Mismatches 131; Indels 197; Gaps 24

QY 19 PPAP-----EAAKKPPPC--HRCRGLVDFKNGMVDTAANKNGGNTAMEEKLSTYESS 71
Db 633 PPAPODVDECARSPPTCTYGRCENTEGSFQCVCPMGFPNMAAGS----- 678
QY 72 EIRLEIEIGLCSSSFEENQMLAEQEHLEAMWLDLAKSEYFDLFFMFCYKTLKVCSSPG 131
Db 679 -----CEDVD-EC-----ENHLACPGQE-----CVN 701
QY 132 TYGPCLACGG-----SQRPSGSGHCSGDSRGDSCSRHMGYQGP 175
Db 702 SF--QCRACPSHHLRGACTVDDECSGAPFCGPHGCT--NTGSRFCSCAPETRAP 756
QY 176 L-----CTD--CMDGYF-----SSLRNETHSICTA-----CDESC 204
Db 757 SGRPGCADVNCLEGGDFPFGHECLNTDGSFACTCAPGYRPGPRGASCLVDDECSDDL 816
QY 205 TCSGL-TNRD-----CG----- 215
Db 817 COSGICITNDGSEFECICPGHRAGBDLASCLVDDECRGPAICSGORCENSPGSYRVR 876

```

QY	216	EC:IGW-V:IDEGACVVDDECAAPPCSAAPFCNKANGSYTC-E:ECDS:SCVGCCTGEGPGN	273
Db	877	DCDDPGYHAGPGCTDDVDCEQYGEYGPETICGAGQRCENMPGSTRCTPPACDPGYQTPPG--GG	933
QY	274	CKE-----CISGY--AREHGCAVDYDECSLEAKTCVKN	305
Db	934	CQVDDEGRNRSFGCAHVAVCNLPGRSFQCLDDQYEGARQGRHCYVDVNEETLQYVC--GA	991
QY	306	ENCYNTPGSYVCVCPDGFETE--DACVPAPAEATEGESPTQLPS	349
Db	992	ALCENNVESGSEFLCVCVNSPEEFEDPMTGRCVPPRTSAGMFGPSGPOAPA	1038
RESULT	14		
ID	Q9BI05	PRELIMINARY;	PRT; 2189 AA.
AC	Q9BI05;		OA.
DT	01-JUN-2001 (TReMBLrel. 17, Created)		
DT	01-JUN-2001 (TReMBLrel. 17, Last sequence update)		
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE	MICRONEME PROTEIN 4.		
GN	MIC4.		
OS	Elmeria tenella.		
OC	Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;		
CC	Eimeria.		
OX	NCBI_TaxID=5802;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-HOUGHTON;		
RA	Tomley F.M., Billington K.J., Bumstead J.M., Clark J.D., Monaghan P.,		
RT	"Emic4: A microneme protein from Eimeria tenella that contains tandem		
RT	arrays of epidermal growth factor-like and thrombospondin type-I		
RL	repeats."		
DR	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ306453; CAC34726.1; -		
DR	HSP: P35555; 1EMN		
DR	InterPro: IPR000152; ASx_hydroxyl.		
DR	InterPro: IPR000561; EGF-like.		
DR	InterPro: IPR001881; EGF_Ca.		
DR	InterPro: IPR000884; TSP1.		
DR	Pfam: PF00008; EGF_24.		
DR	SMART: SM00181; EGF_30.		
DR	SMART: SM00179; EGF_Ca; 30.		
DR	PROSITE: PS00010; ASX_HYDROXYL; 22.		
DR	PROSITE: PS01186; EGF_2; 18.		
DR	PROSITE: PS01187; EGF_Ca; 18.		
DR	PROSITE: PS50092; TSP1; 1.		
KW	Calcium binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat		
SO	SEQUENCE	2189 AA; 224388 MW; C372A420B94FCB2D CRC64;	

	Query Match	Score 271;	DB 5;	Length 2189;
	Best Local Similarity	27.6%;	Pred. No. 5e-17;	
	Matches	92;	Conservative	24; Mismatches 87; Indels 139; Gaps 20;
Oy	120	CVKTL--KVCSPGYTGPDLCAOGGSGRPSGCNGHC-----SGDGR-----QGDG	164	
Dd	273	CNTLTTFECICLDAGDG-----AGTHSPCVDIIDCEKREKPSNCCNNNAVCTNTEGSY	326	
Oy	165	SCRCMHGVGG-----PLCTD--C---MDGFSSLRNETHSICT-----NC	199	
Dd	327	TCACKEFGFSGEGFGAAGCADVDDECANSPCDALASCAINTBSYVTCITNPGEFASSDGHAC	386	
Oy	200	---DE---SCKTCSGLTNRD---GCECEYGVNLDEGACVDVDECAAEPPPCSAOAF	246	
Dd	387	KDVDCACAGTAGACHVSADQCVNDGYSYC-HCLEGFPIGDGVSDVDECAAEAFSPCCANTH	445	
Oy	247	CKNANGSYTCE-----EC-----DSSCYGTGE-----GFG---	272	
Dd	446	CLNTIGSYCECKDKDYGHNMGNAACSIDDCSEASTETPENCNCVNTEGFSFLAKRGYTEL	505	
Oy	273	-----NCKE-----CISGYARE---HGQCADVDECSLAER	299	

[illegible][illegible]

Db 2351 DECSSTGHMCSQHADCKNTMGSYRCLCKDGYTGDTCTDDECSSEMLNLCGNGCULNAP 2410  
 QY 268 G-----EGP-----GNCK-----ECISYA--REHGQ 287  
 Db 2411 GGYRCECDMGFVPSADKACEDIDECSLPNICVFGTCHNLPLGRCEDEIGYELDRSGN 2470  
 QY 288 CADVDECSLAECTVCRKNECNYTPGSGVYCVPGDFE--ETEDACV 331  
 Db 2471 CTDVNEC-LDPTTCI--SGNCVNTPGSYTCDCCPPDFELNPRVGV 2513

## RESULT 16

Q14767 PRELIMINARY; PRT: 1821 AA.  
 AC 014767; 099907;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE LATENT TRANSFORMING GROWTH FACTOR-BETA-BINDING PROTEIN-2 (LTBP-2).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE-95096101; PubMed-7798248;  
 RA Moren A., Olafsson A., Stenman G., Sahlin P., Kanazaki T.,  
 RA Claesson-Welsh L., ten Dijke P., Miyazono K., Heldin C.;  
 RT "Identification and characterization of LTBP-2, a novel latent  
 RT transforming growth factor-beta-binding protein.";  
 RL J. Biol. Chem. 269:32469-32478(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96252879; PubMed-8697098;  
 RA Bashir M.M., Han M.D., Abrams W.R., Tucker T., Ma R.I., Gidson M.,  
 RA Ritty T., Mechem R., Rosenbloom J.;  
 RT "Analysis of the human gene encoding latent transforming growth  
 RT factor-beta-binding protein-2.";  
 RL Int. J. Biochem. Cell Biol. 28:531-542(1996).  
 DR EMBL: Z37976; CA86030.1; -;  
 DR EMBL: S82451; AAB37459.1; -;  
 DR HSSP: P35555; IEMN.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-like.  
 DR InterPro: IPR002212; TB.  
 DR Pfam: PF00683; TB; 4.  
 DR Pfam: PF00683; TB; 4.  
 DR SMART: SM00179; EGF\_CA: 15.  
 DR SMART: SM00179; EGF\_CA: 15.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 13.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; 11.  
 DR PROSITE: PS01187; EGF\_CA: 16.  
 KM Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 FT CONFLICT 897 897 Q -> K (IN REF. 2).  
 FT CONFLICT 1615 1615 O -> E (IN REF. 2).  
 SO SEQUENCE 1821 AA; 195063 MW; C8194FA09D07B94 CRC64;

Query Match 13.4%; Score 268; DB 4; Length 1821;  
 Best Local Similarity 25.4%; Pred. No. 7.9e-17;  
 Matches 90; Conservative 24; Mismatches 94; Indels 146; Gaps 19;

QY 113 PDLEMFVCKTLKVCSPG-----TYGPDLACQGGSGRPSGNGHSGSDGSRQG 162  
 Db 865 PDGYR--CV-----CSPGQLHPSQAYCTDNECL-----RDPGCGKGRCI--NRVG 907  
 QY 163 DGSRCRHNGY-----OGPLCT-----DCMGYFSSSLANE 191  
 Db 908 SYSCFCYPTLATSGATQECODINECEQPCVSGGGGCTNTGSHYHCECDGTYMVRKGH 967

QY 192 THSI-----CTACDESCCTCSG-----LTNRDC----- 215  
 Db 968 QODINECRHPTCTCDGRCVNSPGSYTCLACEBGRGSGSCVDVNECLTFPGVCHAGKCTN 1027  
 QY 216 -----ECEYGVWL--DEGACVDVDECAEPSCAAQFCCKNANGSYTCECCSS----- 262  
 Db 1028 LEGSFRCSCEQGYEVTSDKEKCDQVDECAAR--ASCPTG--LCLTNIEGSPACSAENGWYVN 1085  
 QY 263 -----CVG--CTG--EGPGNCKECISYAE--HGCCADVDECSLAERT 300  
 Db 1086 EDGTACEDLDACAPPGVCPSPVCNTAGSFSCKDGDGGRPSPLGDSCEDEVDECEDPQS 1145  
 QY 301 CVRKNECNYTPGSGVYCVPGDFEET-----EDACVPAPAEATG 341  
 Db 1146 CL--GGECKNTVGSYQCLCPGFGOLANTGVCEVDYNECGEHCAPRHECLNSHG 1197

## RESULT 17

Q089103 PRELIMINARY; PRT: 644 AA.  
 AC 089103;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE C1Q/MBL/SPA RECEPTOR C1QRP.  
 GN IY68 OR C1QRP OR AA4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SVEV;  
 RA Tenner A.J., Kim T.S.;  
 RT "Identification of the mouse genomic DNA for C1qrp.";  
 RN Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA Petrenko O., Lemischka I.R.;  
 RT "Molecular characterization of AA4, an early marker of hematopoietic  
 RT development.";  
 RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RA MEDLINE-99359842; PubMed-10430665;  
 RA Norworthy P.J., Taylor P.R., Walport M.J., Botto M.;  
 RT "Cloning of the mouse homolog of the 126-kDa human C1q/MBL/SP-A  
 RT receptor, C1qrp.";  
 RL Mamm. Genome 10:789-793(1999).  
 DR EMBL: AF074856; AAC63274.1; -;  
 DR EMBL: AF081789; AAC62649.1; -;  
 DR EMBL: AF099939; AAD47906.1; -;  
 DR EMBL: AF099938; AAD47906.1; JOINED.  
 DR HSSP: P35555; IEMN.  
 DR MGD: MGI:106664; Iy68.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_CA.  
 DR InterPro: IPR001304; lectin\_c.  
 DR Pfam: PF00008; EGF; 5.  
 DR Pfam: PF00059; lectin\_c; 1.  
 DR SMART: SM00034; CLECT; 1.  
 DR SMART: SM00179; EGF\_CA: 3.  
 DR SMART: SM00001; EGF\_CA: 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE: PS00041; C-TYPE LECTIN\_2; 1.  
 DR PROSITE: PS01186; EGF\_2; 3.  
 DR PROSITE: PS01187; EGF\_CA; 3.  
 KM Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;  
 KW Receptor; Repeat.  
 SO SEQUENCE 644 AA; 69354 MW; EB4351648BF8635A CRC64;

Query Match 13.3%; Score 266.5; DB 11; Length 644;  
 Best Local Similarity 27.0%; Pred. No. 3.4e-17;  
 Matches 11; Conservative 40; Mismatches 119; Indels 141; Gaps 28;

QY 12 LPLLLLPAPPAEAKK--ETPC-----HRCGLVDKRN-QGMVDTAKKNRGGGNTAME 61  
 DB 143 LILDLSTPHPSHLRPMHSPGTPPEAPGNSIEGFLCKRFNFKMCRPLALG-GPGVTTYT 201  
 QY 62 ---EKLTSYSESEIRLEILEGLC-----ESSDEPCNOMLEAOREHLEAMWLQKSEY 112  
 DB 202 TTPQATSTSLA--VPFASVANACGDEAKSETHYFLCN-----KT 241  
 QY 113 PDLFEWFCVTKLVKCCSPETYPGDLACGGSGQRPCS-GNGHCSGGSGRQDGS--CRCH 169  
 DB 242 PGIFHW-----SSGFLVSPKFG-----CSFNNGCGQDCCFEGDGSFPGCGR 285  
 QY 170 MGYS-----GPLCTD-----CMDGTFSSLRNTHST--- 195  
 DB 286 PGRLLDLVTCASRNPCSNPCTGGMCHSVPLSENHYTCRCPGYOLD-SSQVHCVDID 344  
 QY 196 -C--TACDSCKTKSGLTNRDCECEVWYLD---EGACVDVECAEAPPPCSAAGFCNK 249  
 DB 345 ECODSPCADQCVNTLG--SFHC-ECWVGTOPSGPKKEACEDVDECAANSPC--AAGCIN 399  
 QY 250 ANGSYTCECDSSCVGCTGEBGPNCECISGY---AREHQCADVDECSLAE-KTCVRKN 305  
 DB 400 TDSFYC-----SCKE---GYIVSGEDSTQCEDIDECSDARGNPC--D 437  
 QY 306 ENCYNTPGYSVCVCPDGFETETACVPPAFAETEG-----ESTQLPKRED 352  
 DB 438 SLCEFTDGSFRCGCPGWE-----LAPNGVFCSGTVESELPARPORKED 482

RESULT 18  
 096J98 PRELIMINARY; PRT; 2809 AA.  
 AC 096J98;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE FIBRILLIN3.  
 GN KIAA1776.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRIN;  
 RX MEDLINE=21245130; PubMed=11347906;  
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 8:85-95(2001).  
 DR EMBL; AB053450; BAB47408.1;  
 SQ SEQUENCE 2809 AA; 300323 MW; 20C04CC006C0161F CRC64;

Query Match 13.2%; Score 264.5; DB 4; Length 2809;  
 Best Local Similarity 23.4%; Pred. No. 2.8e-16;  
 Matches 112; Conservative 37; Mismatches 146; Indels 183; Gaps 25;

QY 10 GLPLLLLPAPPAEAKKPTPCRCRGLVDKRNQGMVDTAKKNRGGGNTAME-----KT 64  
 DB 2239 GMP-----LPGSGEGCTDDNECAQPDLC--VNGRCVNTA-----GSFRCDCCDGGFOPSPFT 2288  
 QY 65 LSKYSESEIR-----LLEILEGLCSGSDPFCNOMLEAOREHLEAMWLQKSEYDILEWFC 120  
 DB 2289 LT--ECHDIRGCPFAEVLQTMKRS-----LSSSE-----A 2318  
 QY 121 VRTLWCCSPGT-YGPDCLAC-----OGGSORPCSGNCHGSGDG- 158

DB 2319 VYRAECCCGGROWGPRCLCPGTSAAYRKLCPHSGSYTAGRDVDECRMLAHLCANGE 2378  
 QY 159 --SRODGSRCRMGX--QGPLCTDCMDGYESSLRNTHSTICTACDESKCT-----CS-- 207  
 DB 2379 CINSLSFRCCHQAGATTPTATITCID-----MDECSQVPRPCTFLCKNTGSEFLCSCP 2432  
 QY 208 -----GLTNRDCE-----CEVWYLDGACVYDVECAAE 237  
 DB 2433 RGYLEEDRTCKDDECTSRQHNQCFLCVNTVGAFTCTCPGPFQHHQACFDNDGCSAO 2492  
 QY 238 PPPCSAOFCKNANGSYTEBED-----SSCVGCTG---EGPNCK----- 275  
 DB 2493 PGPCGAGHCHHTPGSFRC-ECHQGFTLVSSHGCEDEVCEGPHRCQGGCONQLOGYRC 2551  
 QY 276 EICSGYAR--EHQGCADVDECSLAEKTCVRKNENCTNTPGYSVCVCPDGE-----E 325  
 DB 2552 SCPQGTQSHQMAQVDENECALSPETC--GSAKCNITLIGFRVCVPSGFDYDQALGGCQ 2609  
 QY 326 TEDACV-----PPAEAETEG-----SPTQLPKREDL 353  
 DB 2610 DVDECAGRGRCPSYSCANTPGFTLGCQPGYFRAGGHCVSGLGSPGPDTPDKREL 2667

RESULT 19  
 028019 PRELIMINARY; PRT; 1963 AA.  
 AC 028019;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE LATENT TGF-BETA BINDING PROTEIN-2.  
 GN LTB-2.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96069403; PubMed=8524260;  
 RA Gibson M.A., Hatzinikolas G., Davis E.C., Baker E., Sutherland G.R.,  
 RA Mechnam R.P.;  
 RT "Bovine latent transforming growth factor beta 1-binding protein 2:  
 RT molecular cloning, identification of tissue isoforms, and  
 RT immunolocalization to elastin-associated microfibrils.";  
 RL MOL. Cell. Biol. 15:6932-6942(1995).  
 DR EMBL; U35363; AAA91455.1;  
 DR HSSP; P35355; IEMN.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF-Ca.  
 DR InterPro; IPR002212; TB.  
 DR Pfam; PF00008; EGF; 16.  
 DR Pfam; PF00683; TB; 4.  
 DR SMART; SM00179; EGF\_CA; 15.  
 DR SMART; SM00001; EGF-like; 5.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 13.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN2.  
 DR PROSITE; PS01186; EGF\_2; 11.  
 DR PROSITE; PS01187; EGF\_CA; 16.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 KW SEQUENCE 1963 AA; 211511 MW; E8E45F47FAF3B904 CRC64;

Query Match 13.2%; Score 264; DB 6; Length 1963;  
 Best Local Similarity 26.2%; Pred. No. 2.1e-16;  
 Matches 85; Conservative 21; Mismatches 84; Indels 134; Gaps 17;

QY 128 CSPG-----TYGPCLACQGGSGRPGSGNGHCSGDSRQDGSRCRMGX----- 172  
 DB 985 CSPGRLRHPQAYCTDDNECL-----RDPCKGRGNCV--NRVGSYSFCYIPGKILATS 1035

```

QY 173 -----OgPLCTDCMDYFSSLRNETHSI----- 195
Db 1036 GATGECQDIDCEQDGVGYSRRCNTNTGSHYCHCEDQGYIMWRKHCODINCRHPTCPD 1095
QY 196 -----CTACDE-----SC-----KTCGSLNRDCGEEVAVL 223
Db 1096 GKCVNSRGSYCTLCEEBYRGCGSCVDVNECLTPGYCTHCTCINTLESGFRCSGQYEV 1155
QY 224 --DEGACVDVDECAAEPPPCSAOFCRKNANGSYNCEBDS----- 262
Db 1156 TPDKGCKNDVDECAIR-ASCTG-LCLNTBGSFTCSACESYWNEDGTACEDDDECAEP 1213
QY 263 --CVG--CTG-EGPGNKECISGY---AREHQCADVDECSLAETCYRKNENCYNTPGS 314
Db 1214 GVCSSGYCTNTAGSFSCRDCAGYPSALGH-TCEDVDECDPSSCL--GGECKNTAGS 1270
QY 315 YVCYCPDGF-----ETEDACV 331
Db 1271 YQCLCPGPFOLANGTCVDECDV 1294

RESULT 20
ID P87363 PRELIMINARY; PRT; 708 AA.
AC P87363;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE FIBRILLIN-1 (FRAGMENT).
GN FBN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxId=9031;
RN [1]
SEQUENCE FROM N.A.
RA Zhou G., Price C.E., Rosenquist T.H., Gadsen P.F., Godfrey M.,
RT "Partial cloning and sequencing of chick fibrillin-1 cDNA."
RL In Vitro Cell. Dev. Biol. Anim. 36:19-25(2000).
DR EMBL: U08872; AAB48531.1; -.
DR HSSP: P07204; 2ADX.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002212; TB.
DR InterPro: IPR001491; Thrombomodulin.
DR Pfam: PF00008; EGF_13.
DR PRINTS: PR00907; THROMBOMODULN.
DR SMART: SM00179; EGF_CA_14.
DR PROSITE: PS00010; ASX_HYDROXYL; 13.
DR PROSITE: PS01186; EGF_2; 10.
DR PROSITE: PS01187; EGF_CA; 13.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
FT NON_TER 708
FT SEQUENCE 708 AA; 76164 MW; C247271CIDF73361 CRC64;

Query Match 13.1%; Score 263; DB 13; Length 708;
Best Local Similarity 26.5%; Pred. No. 8.1e-17;
Matches 93; Conservative 28; Mismatches 128; Indels 102; Gaps 18;

```

```

Db 285 RCLCYDGFMASEDKKTCVDVNECDLHPNICALSGTCENTKSGFICHCMDGSGKGGTTGCT 344
QY 179 DCMGYSFSSLRNETHSITCTACDESCRTCSGLTNRDCGEEVAVLDEGACVDVDECAEP 238
Db 345 DINECEIAHNCDBAHVATNIPGSFK-CS-----CSSGNGICNGIKCYDLDCECSNGT 394
QY 229 PCSAOFCKRKNANGSYNCEBDSGCVGTGEGPGNCKRCISGYAREHQCADVDECSLA 298
Db 395 HKCSPHADCKNTMGSYRC-----LCKE--GYGDGFTCTDLDCECSNLT 435
QY 299 KTCVRKNENCYNTPGSYVCVCPDGFEEEDACVPAAEATEGSSPQLPS 349
Db 436 NLC--ENQGCLNAPGCGYRCBDMGF-----LPSLOKACEDDDECSLTPN 477

RESULT 21
ID O35806 PRELIMINARY; PRT; 1764 AA.
AC O35806;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE LATENT TGF-BETA BINDING PROTEIN-2 LIKE PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
SEQUENCE FROM N.A.
RA Kriehn K.;
RC TISSUE=LUNG;
RT "TGF-dependent expression of a rat LTBP-2 homologue in astrocytes and
RT C6 glioma cells."
RL Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: Y12760; CA73300.1; -.
DR HSSP: P35555; 1EMN.
DR InterPro: IPR002086; Aldehyde dehydr.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002212; TB.
DR Pfam: PF00008; EGF_16.
DR Pfam: PF00683; TB; 4.
DR SMART: SM00179; EGF_CA_14.
DR PROSITE: PS00010; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00010; ASX_HYDROXYL; 12.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 11.
DR PROSITE: PS01187; EGF_CA; 15.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
FT SEQUENCE 1764 AA; 189866 MW; 245D57F9CE3386D0 CRC64;

Query Match 13.1%; Score 262.5; DB 11; Length 1764;
Best Local Similarity 26.0%; Pred. No. 2.6e-16;
Matches 87; Conservative 21; Mismatches 91; Indels 135; Gaps 17;

```



DB 1022 FRCSCEPEYELAPDKKGRDVEDCASR-ASCP7G-LCLNTEGSEFTGACOSGYWVEDGT 1079  
 OY 262 SCVG-----CTGE-GPGNCKEVCISGYARE--HGOCADYDECSLAETKCYRK 304  
 DB 1080 ACEDDDECAFPYGVPCVCTNTVGSFSCDKCDGFRPSPGLGNSCEDVEDCEGQNSCL-- 1137  
 OY 305 NENCYNTPGSYVCVCPDGE-----ETEDACV 331  
 DB 1138 GGECKNTDGSYQCLCPQGFQLANGTWCEDVDECV 1171

## RESULT 22

O9WU8 PRELIMINARY; PRT; 2872 AA.  
 ID 09WU8  
 AC 09WU8;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE FIBRILIN-1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99032689; PubMed=9815129;  
 RA Kanwar Y.S., Ota K., Yang Q., Kumar A., Wada J., Kashihara N.,  
 RA Peterson D.R.;  
 RT "Isolation of rat fibrillin-1 cDNA and its relevance in metabolic  
 RT development.";  
 RL Am. J. Physiol. 275:F710-F723(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kanwar Y.S.;  
 RT Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.  
 RL EMBL: A133059; AAD34438.1; -.  
 DR HSSP; P35355; IAPU.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR002212; TB.  
 DR InterPro: IPR000822; ZnF-C2H2.  
 DR Pfam; PF00008; EGF; 45.  
 DR Pfam; PF00683; TB; 9.  
 DR SMART; SM00179; EGF\_CA; 41.  
 DR SMART; SM00001; EGF-like; 5.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 42.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE; PS01187; EGF\_2; 38.  
 DR PROSITE; PS01187; EGF\_CA; 41.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_1.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SO SEQUENCE 2872 AA; 312069 MW; 0C4FF3B87A80280 CRC64;

Query Match 13.1%; Score 262.5; DB 11; Length 2872;  
 Best Local Similarity 25.6%; Pred. No. 4.5e-16;  
 Matches 104; Conservative 29; Mismatches 123; Indels 151; Gaps 23;

OY 30 PCHRCGLVDK---NCGVNDPAKKNPGGNTAMEKILSKESSEIRLLELGLCES 85  
 DB 1168 PHCRVNLIGKYECACNPGYHPHDLRF-----CYDIECSITMNGCET 1211  
 OY 86 -----SDFCN-----OMLEAOEHLAWMLQLKSEYDPLFEMFCVTKLVC-----CS- 129  
 DB 1212 FCTNSGSEYSCQPGFALMPDR-----SCDIDE--CEDNPNITDGGQCTN 1257  
 OY 130 -PGYAPDLACGG---GSQRPCSGNGHC-----SGD-GSRQDGSRCRMGYQGP 175  
 DB 1258 IPGEY--RCICYDGFMASEDMKTCVDVNECDLNPICLSTCENTRGFSFCHCDMGYSGR 1315  
 OY 176 L-----CTDCMDGTFSSLRNETHSICACDESCCTCGSLNRDCGCEVGVNIDEGACVDV 231  
 DB 176 L-----CTDCMDGTFSSLRNETHSICACDESCCTCGSLNRDCGCEVGVNIDEGACVDV 231

DB 1316 KKGCTGTDINECEIGHANCGRNAVCTNAGSFK-CS-----CSPGWIGDGIKCTDL 1365  
 OY 232 DECAAPPPCSAOPCKNANGSYTC-----EEDSSCGCTGEEP----- 271  
 DB 1366 DECSNETHMCSQHADCKNTMGSTRCLCKRGYTGDEFTCTDDEG--SENINLSGNQCLNA 1424  
 OY 272 -----GNCX-----ECISGYA--REHG 286  
 DB 1425 PAGYRCECDMGFVPSADGACEDINECSLPNICVFCTCHNLPGLFRCECEIGYELDRSGG 1484  
 OY 287 QCADVDECSLAETCYRKNNENCTNPGSYVCVCPDGE--ETEDACV 331  
 DB 1485 NCTDVNEC-LDPTTCL--SGNCVNTPGSYTCDDCPDFELNPTVGCV 1528

## RESULT 23

O75441 PRELIMINARY; PRT; 669 AA.  
 ID 075441  
 AC 075441;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE LATENT TRANSFORMING GROWTH FACTOR-BETA BINDING PROTEIN 4  
 DE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98325059; PubMed=9660815;  
 RA Saharinen J., Taipale J., Monni O., Keski-Oja J.;  
 RT "Identification and characterization of a new latent transforming  
 RT growth factor--binding protein, LTRP-4.";  
 RL J. Biol. Chem. 273:18459-18469(1998).  
 DR EMBL: AF054502; AAC39882.1; -.  
 DR HSSP; P35355; IEMN.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR002212; TB.  
 DR Pfam; PF00008; EGF; 6.  
 DR Pfam; PF00683; TB; 2.  
 DR SMART; SM00179; EGF\_CA; 6.  
 DR SMART; SM00001; EGF-like; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 6.  
 DR PROSITE; PS01187; EGF\_2; UNKNOWN\_2.  
 DR PROSITE; PS01187; EGF\_CA; 8.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 FT NON\_TER 1  
 FT TER 669  
 SO SEQUENCE 669 AA; 71702 MW; 2C76C44BB11428A4 CRC64;

Query Match 13.0%; Score 260.5; DB 4; Length 669;  
 Best Local Similarity 27.3%; Pred. No. 1.3e-16;  
 Matches 89; Conservative 27; Mismatches 99; Indels 111; Gaps 19;

OY 83 CESSDFECNMLEAOEHLAWMLQLKSEYDPLFEMFCVTKLVCCSPGTYPDLACOG 142  
 DB 15 CEDVD-EC-----ENHLACPGQE-----CVN-----SPGSF--QCRACPS 46  
 OY 143 G-----SOPCSGNGHCSGDSRQDGSRCRMGYQGP-----CTD-- 179  
 DB 47 GHHLHGRCTVDDECSGAPPCGPHCHC---NTBSFPGSCAPGRAPSGRGACADVN 103  
 OY 180 -CMDGYF-----SSLRNETHSICTA-----CDBSCCTGSGL--TNRDC 214  
 DB 104 ECLGDFCFPHGECLNTDGSFACTCAPGYRPGASCIDVDECSBEDLCQSGICTNTDG 163  
 OY 215 G-EC-----EVGWVLDDEACVDVDECAAPPPCSAOPCKNANGSYTCEDSSVCGCT 267  
 DB 164 SPECICPGHRRAG--PDILASCLDVDECRERGPALCGSQRCENSPSGYR----- 210



```

QY 128 CSPG-----TYGPDLACGGGSGRPCS-GNGHCSGDSRGDSCRCRMGYO----- 173
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 91 CNBGRYSHPOHRYCYDVNCEA---EPGPRGIGICMNTG---GSYNCHCNRGYRLHVGAG 144
QY 174 GPLCTD-----CMDGYFS-----SLRNETHSICTACDESCCKTCSGL 209
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 145 GASCVDLNECARPHLCGGDGFICINFPGHYKNCNCPGYRLKASRPVCEBIDE-CRDPSSC 203
QY 210 TNRDGCE-----CEYGM-VLDEGACVDYDECAAEPPPCSAOFCNKANGSYTC--- 256
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 204 PGKCKENKPGSFCKIACQPGYRSOGGACRDVNEC-AEGSPCSPG-WCENLPGSFCTCA 261
QY 257 -----EEDCSSCV---GCTGEGRGNCK-ECISGY--AREHGOCADVDECS 295
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 262 QGYAPAPDGRSCLDYDECAADVCDNGICSNTPGSPQCCLSGYHLSRDRSHCEBIDECD 321
QY 296 LAEKTQVRKNECYNTPGSYVCVCPDGF-----EETDCAVPPAEAE---- 337
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 322 F-PACI---GGDCINTNGSYRCLCPQGHRLVGRKCODIDECSDPSLCLPHGACKNLIG 378
QY 338 -----ATEGESPTQ 346
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 379 SYVCVDEGFTPTQ 392

```

## RESULT 27

Q9NS15 PRELIMINARY; PRT: 1242 AA.

```

AC Q9NS15;
ID 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LATENT TRANSFORMING GROWTH FACTOR BETA BINDING PROTEIN 3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Saharinen J., Penttinen C., Keski-Oja J.;
RT "Cloning of human LTRP-3."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF135960; AAF62352.2; -.
DR HSSP; P08709; 1BF9.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR Pfam; PF000683; TB; 4.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00001; EGF_Like; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 10.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 11.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 1242 AA; 133028 MW; 155C87FB69AB221B CRC64;

```

Query Match 12.8%; Score 256.5; DB 4; Length 1242;  
 Best Local Similarity 26.8%; Pred. No. 6.5e-16;  
 Matches 84; Conservative 32; Mismatches 91; Indels 107; Gaps 20;

```

QY 128 CSPG-----TYGPDLACGGGSGRPCS-GNGHCSGDSRGDSCRCRMGYO----- 173
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 587 CNBGRYSHPOHRYCYDVNCEA---EPGPRGIGICMNTG---GSYNCHCNRGYRLHVGAG 640
QY 174 GPLCTD-----CMDGYFS-----SLRNETHSICTACDESCCKTCSGL 209
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 641 GASCVDLNECARPHLCGGDGFICINFPGHYKNCNCPGYRLKASRPVCEBIDE-CRDPSSC 699

```

```

QY 210 TNRDGCE-----CEYGM-VLDEGACVDYDECAAEPPPCSAOFCNKANGSYTC--- 256
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 700 PGKCKENKPGSFCKIACQPGYRSOGGACRDVNEC-AEGSPCSPG-WCENLPGSFCTCA 757
QY 257 -----EEDCSSCV---GCTGEGRGNCK-ECISGY--AREHGOCADVDECS 295
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 758 QGYAPAPDGRSCLDYDECAADVCDNGICSNTPGSPQCCLSGYHLSRDRSHCEBIDECD 817
QY 296 LAEKTQVRKNECYNTPGSYVCVCPDGF-----EETDCAVPPAEAE---- 337
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 818 F-PACI---GGDCINTNGSYRCLCPQGHRLVGRKCODIDECSDPSLCLPHGACKNLIG 874
QY 338 -----ATEGESPTQ 346
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 875 SYVCVDEGFTPTQ 888

```

## RESULT 28

Q9H7K2 PRELIMINARY; PRT: 1382 AA.

```

AC Q9H7K2;
ID 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FLJ00070 PROTEIN (FRAGMENT).
GN FLJ00070.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Thissue=Spleen;
RA Chava O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024477; BAB15767.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR Pfam; PF000683; TB; 4.
DR SMART; SM00181; EGF; 15.
DR SMART; SM00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR EGF-like domain; Glycoprotein; Hydroxylation.
KW NON_TER
SQ SEQUENCE 1382 AA; 147283 MW; AF695AACB9913DF1 CRC64;

```

Query Match 12.8%; Score 256.5; DB 4; Length 1382;  
 Best Local Similarity 26.8%; Pred. No. 7.3e-16;  
 Matches 84; Conservative 32; Mismatches 91; Indels 107; Gaps 20;

```

QY 128 CSPG-----TYGPDLACGGGSGRPCS-GNGHCSGDSRGDSCRCRMGYO----- 173
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 680 CNBGRYSHPOHRYCYDVNCEA---EPGPRGIGICMNTG---GSYNCHCNRGYRLHVGAG 733
QY 174 GPLCTD-----CMDGYFS-----SLRNETHSICTACDESCCKTCSGL 209
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 734 GASCVDLNECARPHLCGGDGFICINFPGHYKNCNCPGYRLKASRPVCEBIDE-CRDPSSC 792
QY 210 TNRDGCE-----CEYGM-VLDEGACVDYDECAAEPPPCSAOFCNKANGSYTC--- 256
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 793 PGKCKENKPGSFCKIACQPGYRSOGGACRDVNEC-AEGSPCSPG-WCENLPGSFCTCA 850
QY 257 -----EEDCSSCV---GCTGEGRGNCK-ECISGY--AREHGOCADVDECS 295
   1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
DB 851 QGYAPAPDGRSCLDYDECAADVCDNGICSNTPGSPQCCLSGYHLSRDRSHCEBIDECD 910

```

QY	296	LAERKAVKNECVNTPCSYCVCGDGF-----EEDBDACVPAEAE----	337
Db	911	F-PAACI--GGGCIITMNSRCLCPQGHRLVGRKKCDIDECSDPSLCLPHGACKNLIG	967
QY	338	-----ATEGESPTQ 346	
Db	968	SYVCVCDEGFPPTQ 981	
RESULT	29		
Q9J126			
Q9J126	PRELIMINARY;	PRT: 643 AA.	
AC	Q9J126:		
DT	01-OCT-2000 (TREMBLrel. 15	Created)	
DT	01-OCT-2000 (TREMBLrel. 15,	Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19,	Last annotation update)	
DE	GIORP.		
GN	GIORP.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MISTER; TISSUE=LUNG;		
RC	MEDLINE=20507883; PubMed=10934210;		
RA	Dean Y.D., McGreal E.P., Akatsu H., Gasque P.;		
RT	"Molecular and Cellular Properties of the Rat AA4 Antigen, a C-type		
RT	lectin-like receptor with Structural Homology to Thrombomodulin."		
RL	J. Biol. Chem. 275:34382-34392(2000).		
DR	EMBL: AF160978; AAF80402.1; -		
DR	HSSP: P35555; IEMN.		
DR	InterPro: IPR000152; ASX_hydroxyl.		
DR	InterPro: IPR000561; EGF-like.		
DR	InterPro: IPR001881; EGF_Ca.		
DR	InterPro: IPR001304; lectin_c.		
DR	SMART: SM00034; CLECT: 1.		
DR	SMART: SM000179; EGF_CA: 3.		
DR	SMART: SM00001; EGF_like: 2.		
DR	PROSITE: PS00010; ASX_HYDROXYL: 3.		
DR	PROSITE: PS50041; C_TYPN_LECTIN_2: 1.		
DR	PROSITE: PS01186; EGF_2: 3.		
DR	PROSITE: PS01187; EGF_CA: 3.		
KW	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.		
SEQUENCE	643 AA: 68780 MW: FB230742E02A4931 CRC64;		
Query Match	12.8%; Score 256; DB 11; Length 643;		
Best Local Similarity	26.7%; Pred. NO.3.4e-16;		
Matches	108; Conservative 40; Mismatches 125; Indels 132; Gaps 28;		
QY	12	LPLLLPPAPPAAK--PTPC-----HRCGLVDKFN-OGWVDYAKKNFGGNTAWE 61	
Db	143	LILDLSTLPHPSHLRKMWHESPCTPDADGNSIEGFLCKFNFGKCSPLALG-GPOLITYT 201	
QY	62	---EKLTKYSESSEIRLLEILEGL-ESSDFPCNMLAEQHEHLAMWILQKSEYRDLPE 117	
Db	202	TPFOATTTSLRA--VPFASVANVVGDEAESNTNYL-----CKETAGVNH 246	
QY	118	WFCVTKLVVCCSPGTGYGPDCLACDGGSGSORPCS-GNGHCSGSDSGRODGS--CRCHMGYO- 173	
Db	247	W-----GSSGPLCVSPKRG---CSFNNGGCGOQDCFEFGDGSFRCGCRGPFLL 290	
QY	174	-----GPLC-----IDCMGDTYSSILRNETHSI----C--T 197	
Db	291	LDDLVTCASRNPCSSNPCTGGMGCHSVPLSENYTCHCRGYOLD--SSQVHCVDIDECDS 349	
QY	198	ACDESKTCIKSGLTINBDCECEVEGYVLD---EGACVVDVECAALRPPKCSAAOFCKNANGSY 254	
Db	350	PCDQDCIMTPG--GFHC-ECWVGYYSSSGSKELACEDYDECTAAYSPC--AQGCTINTDSGF 404	
QY	235	TCBECDSGCVGTGEGPGNKCEKICISGY---AREHGQACADVDECSLAEKTCVKNKNCYNT 311	

[illegible]

```
Db 718 GG-----GACRDVNEC-SEGPPCSPG-WCENLPGSYRCTCAOIGRTPT 758
QY 257 -----ECDSGCV---GCTGEGPNCK-ECISGY--AREHGOCADVDECSLAEK----- 299
Db 759 GRUSCIDVDECEGKVCODGICTNTPTGSPFOCCLSGYHLSDRSCDECDPACICG 818
QY 300 -TCVRKNEN-----CYNTPGSYVCCPDGF 323
Db 819 GDCINTNGSYRCLPLGHRVLVGRKCKKDIDECSDPGCLPLPHACENLQGSYVCCDEGF 878
QY 324 EETED 328
Db 879 TLTD 883

RESULT 31
000274 PRELIMINARY; PRT; 652 AA.
AC 000274;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE C1QR(P).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97199258; PubMed=9047234;
RA Nepomuceno R.R., Henschen-Edman A.H., Burgess W.H., Tenner A.J.;
RT "cDNA cloning and primary structure analysis of C1QR(P), the human
RT C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro.";
RL Immunity 6:119-129(1997).
DR EMBL; U94333; AAB53110.1; -.
DR HSSP; P35355; 1EMN.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR01881; EGF_Ca.
DR InterPro; IPR01304; lectin_c.
DR Pfam; PF00008; EGF_5.
DR SMART; SM00034; Clect; 1.
DR SMART; SM00179; EGF_CA; 3.
DR SMART; SM00001; EGF-like; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS50041; C_Type_Lectin_2; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 652 AA; 68577 MW; B7EAB5FE5714A775 CRC64;
SQ
```

Query Match 12.7%; Score 254.5; DB 4; Length 652;  
Best Local Similarity 30.7%; Pred. No. 4, 8e-16;  
Matches 78; Conservative 27; Mismatches 88; Indels 61; Gaps 15;

```
QY 325 -----ETEDACVPP 333
Db 421 DGTQGDVDECVGP 434

RESULT 32
09ET61 PRELIMINARY; PRT; 643 AA.
ID 09ET61;
AC 09ET61;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE C1Q/MBL/SPA RECEPTOR C1QR PRECURSOR.
GN C1QR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PVG.
RX MEDLINE=20545218; PubMed=11093152;
RA Lovik G., Vaage J.T., Dissen E., Szpirer C., Ryan J.C., Rolstad B.;
RT "Characterization and molecular cloning of rat C1qr, a receptor on NK
RT cells.";
RL Eur. J. Immunol. 30:3355-3362(2000).
DR EMBL; AF136537; AAG01572.1; -.
DR HSSP; P35555; 1EMN.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR01881; EGF_Ca.
DR InterPro; IPR01304; lectin_c.
DR Pfam; PF00008; EGF_5.
DR SMART; SM00034; Clect; 1.
DR SMART; SM00181; EGF_5.
DR SMART; SM00179; EGF_CA; 5.
DR SMART; SM00001; EGF-like; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS50041; C_Type_Lectin_2; 1.
KW Signal; Receptor.
FT SIGNAL 1 23 POTENTIAL.
SQ SEQUENCE 643 AA; 68781 MW; 9AE4C93AD943DB6 CRC64;
SQ
```

Query Match 12.7%; Score 254; DB 11; Length 643;  
Best Local Similarity 26.7%; Pred. No. 5, 3e-16;  
Matches 108; Conservative 39; Mismatches 126; Indels 132; Gaps 28;

```
QY 12 LPLLPLPPAPAAK--PTPC-----HRCGLVDFKN-OGVYDTAKKFGGNTAME 61
Db 143 LILDLSTLKHPPHLLPKMHESPCTPDAPGNSIEGFLCKNFKMGCSPLALG-GPQLVYT 201
QY 62 -----EKLTKYESSEIRLEILLEGIC-ESSDFPCNOMLEAHEHLAMWILQKSEYRDLFE 117
Db 202 TPPOATSSLSKA--VPFASVANVVCGDDESKTNYL-----CETTAGVFN 246
QY 118 WFCVATLKWCCSPGYGPDCLACOGGSGORPCS-GNGHCSGDSRGSDGS--CRCHNGYO- 173
Db 247 W-----GSSGPLCYSPKFG-----CSFNNGCCQGDPEEGSDGSFRGCGCRPRL 290
QY 174 -----GPLC-----TDCKMDGYFSSLRNETHSI--C--T 197
Db 291 LBDLVTCASRNPCSSNPTGGMCHSVPLSENYTCHCPRGYLD--SSQVHCYVIDECEDS 349
QY 198 ACDESCKTSGLTNRDCGCEVGVWLD--EGACYVDVDECAAEPPPCSAAGOCCKANASY 254
Db 350 PCDOECINTPG--GFRHC--ECWVGYQSSGSKFEACVEDECTAAYSPC--AOGCTWTDSF 404
QY 255 TCEEDDSSCVGCTGEGPNCKECISGY--AREHGOCADVDECSLAEKTCVARKNENCYWT 311
Db 405 YC-----SKCE--GYIWSGEDSTQCEDIDEC--LGNPC--DTLCTWT 440
```

```

Oy      312 PSYSCVCPDGEFTEDACVPAAEADEG-----ESPQLPERD 352
          || | | | | | | | | | | | | | | | | | | | |
Db      441 DGSFRCGPAGE-----LAPNGVSTRGSMFSLPARPAPKED 479

RESULT# 33
ID       Q23587 PRELIMINARY; PRT: 3507 AA.
AC       Q23587;
DT       01-NOV-1996 (TREMBLrel. 01, Created)
DT       01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE       01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE       ZK783.1 PROTEIN.
CN       ZK783.1..
OS       Caenorhabditis elegans.
OC       Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC       Rhabdilitidae; Peloderinae; Caenorhabditis.
RX       NCBI_TaxID=6239;
RN       [1]
RP       SEQUENCE FROM N.A.
RC       STRAIN-BRISTOL NZ;
RX       MEDLINE=94150718; PubMed=7906398;
RA       Wilson R., Alnsough R., Anderson K., Baynes C., Berke M.,
RA       Bonfield J., Burton G., Connell M., Copsey T., Cooper J., Coulson A.,
RA       Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA       Gardner A., Green P., Hawkins T., Kirsten J., Laister N., Latreille P.,
RA       Jones M., Keshav J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA       Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA       Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA       Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA       Watson A., Weinstock L., Wilkinson-Sprat J., Wohldman P.;
RT       "2. Mb of contiguous nucleotide sequence from chromosome III of C.
RT       elegans."
RL       Nature 368:32-38(1994).
RN       [2]
RP       SEQUENCE FROM N.A.
RC       STRAIN-BRISTOL NZ;
RA       Favello A., Vaudin M.;
RT       "The sequence of C. elegans cosmid ZK783."
RL       Submitted (NOV-1994) to the EMBL/genbank/DBJ databases.
RN       [3]
RP       SEQUENCE FROM N.A.
RC       STRAIN-BRISTOL NZ;
RA       Waterston R.;
RL       Submitted (JUN-1998) to the EMBL/genbank/DBJ databases.
DR       EMBL: U13646; AAC24418.1; -.
DR       HSSP: P00736; IARO.
DR       InterPro: IPR000152; Asx_hydroxyl.
DR       InterPro: IPR000561; EGF-like.
DR       InterPro: IPR001881; EGF-Ca.
DR       InterPro: IPR001254; Trypsin.
DR       InterPro: IPR001507; zona_pellucida.
DR       Pfam: PF00008; EGF_15.
DR       SMART: SM00179; EGF_CA_12.
DR       SMART: SM00001; EGF_like_16.
DR       SMART: SM00241; ZP_1.
DR       ProSITE: PS00010; ASX_HYDROXYL_9.
DR       ProSITE: PS00022; EGF_1; UNKNOWN_1.
DR       ProSITE: PS01186; EGF_2_22.
DR       ProSITE: PS01187; EGF_CA_13.
DR       ProSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
KW       Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ       SEQUENCE 3507 AA; 365315 MW; 154F0B68787AD9DF CRC4;

Query Match      12.6%; Score 253; DB 5; Length 3507;
Best Local Similarity 19.7%; Pred. No. 4.6e-15;
Matches 103; Conservative 41; Mismatches 100; Indels 280; Gaps 23;

```

Oy	126	-VCC-----	SPGTGPPC-----	LACOGSGORRC-----	SGNGCSG-----	156	
Db	1501	STCCGANAKCVNRKGT	SCCEBNGFLGDGVCYPTTKR	PCDSTQSSKSHCS	SNNSCEVD	1560	
Oy	157	-DGS-----					
Db	1561	TVDSVEBCKECMGYK	KKSGVCECIDNECAVEKAPCS	SLNANCVMNMGTF	SCSCQKQYRBDG	1620	
Oy	176	-LCID-----	C-MDGYFSLNRETHS-----	ICT-----	ACDESK	204	
Db	1621	FMCDINDECEDERHPC	HAPHAECTNLEG-----	SFKCECHSGEGDG	IKCTNPLERSCDEVK	1677	
Oy	205	TCSGLTJNRDC-----	GCEGVGNVDE--	GACVDVDECAEAPP	PPCS-AAORCK	248	
Db	1678	FCGRVDHVSCLSVAT	YTGSLSSVCECEPEGRF	REKESNSCYDIDEC	EBESRNNCDPASANCV	1737	
Oy	249	NANGSYTCE-----	ECDSGVC-----		TGCGP	271	
Db	1738	NTGEGYRCECAEGY	EGEGVCTDIDECDRMAC	CDMSAMCINRMGSCG	CKMAGYTGDDGA	1797	
Oy	272				GKCK-----	275	
Db	1798	TCIKIEEPKSDKTACT	DEMRLCELEKKOCTVDEBEV	POGACCLPGHHP	INGTQSLQI	1857	
Oy	276		ECI-----	SGYAREHGOCAD	DECSLAEKTCVR	KNENCY	309
Db	1858	SGLCQAKNRCKNAK	HECIDHPDHSFPCSPRG	FGIDMCTDDVDECB	NA-GMCDENTNCE	1916	
Oy	310	NTGPSYVCVCPDGE	FETEDACVPPAEAEAT	EGESPQLPSRDL	353		
Db	1917	NTIGSFNCVCEG	EGFKKVDCKCYVDEK	Q-----	PNREKI	1950	
RESULT	34						
O9VR08		PRELIMINARY;		PRT;	3680	AA.	
AC	O9VR08						
DF	01-MAY-2000	(TREMBlrel. 13, Created)					
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)					
DT	01-DEC-2001	(TREMBlrel. 19, Last annotation update)					
DE	CG15637	PROTEIN.					
GN	DP OR CG15637.						
OS	Drosophila melanogaster (Fruit fly).						
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;						
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;						
OC	Ephydroidea; Drosophilidae; Drosophila.						
OX	NCBI_TaxId=7227;						
RN							
RP	SEQUENCE FROM N.A.						
RC	STRAIN=BERKELEY.						
RX	MEDLINE-20196006; PubMed-10731132;						
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,						
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,						
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,						
RA	Sutton G.G., Wortman J.R., Yeakley M.D., Zhang O., Chen L.X.,						
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,						
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,						
RA	Abdill J.F., Agbayanl A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,						
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,						
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,						
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,						
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,						
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,						
RA	de Pablos B., Delcher A., Deng Z., Duan A.D., Dew I., Dietz S.M.,						
RA	Dodson K., Doup L.E., Downes M., Mayas-Rocha S., Dunkov B.C., Dunn P.,						
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferliera S., Fleischmann W.,						
RA	Foster C., Gabrielian A.E., Galy N.S., Gelbart W.M., Glaeser K.,						
RA	Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,						
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck K.,						
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,						
RA	Jatani M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,						
RA	Kimmel B.E., Kolira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,						

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shie B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003576; AAF5100.1; -.  
 DR HSSP: P35555; IEMN.  
 DR Flybase: FBgn0000488; dp.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR002899; EB.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR002603; ET.  
 DR Pfam: PF00008; EGF\_24.  
 DR Pfam: PF00008; EGF\_24.  
 DR Prodom: PD003573; ET; 1.  
 DR SMART: SM00179; EGF\_CA; 16.  
 DR SMART: SM00001; EGF\_Like; 35.  
 DR SMART: SM00274; FOLN; 5.  
 DR SMART: SM00289; WRI; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 18.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; 18.  
 DR PROSITE: PS01187; EGF\_CA; 17.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 KW  
 FT  
 SQ SEQUENCE 3680 AA; 388637 MW; E5E972E1A3479EFF CRC64;

Query Match 12.6%; Score 253; DB 5; Length 3680;  
 Best Local Similarity 24.6%; Pred No. 4.9e-15;  
 Matches 102; Conservative 29; Mismatches 98; Indels 186; Gaps 25;

QY 82 LCESDDEFCNOMLEAOEHLLEAMWQ---LKSEYDLEFEMFVKTLKYC-----C-SP 130  
 DB 383 LC-SNFDCTNNAECIEN--QCFCLDGFEPFGISSCVYDIDE--CRTHAEVCGPHAQCLNTP 437  
 QY 131 GYYGPDCLACQGS-----QRP-----CSGNHCSGDSRGDSCRCRHMGT----- 172  
 DB 438 GSYGCEGCEAGYVGSPPRMACKQPCEDVRCGAHAAYCKPD--ONEAYCVGEDGWTNPSDV 494  
 QY 173 -----OGPLCTDCMDGYFSS-----LRNETHSITACDESKT-- 205  
 DB 495 AAGCVYDIDCDVWHGPRGSGCONATCTNSAGFTCACPFGFSDDPISKCVYDVE--CRIGA 553  
 QY 206 -----CSGLT-----NRDC----- 214  
 DB 554 SKGAGAEVCVNVPGGYYTCRCPGNTIADDPVSRCVPIVSCSANEDCPGNSICDATKRCL 613  
 QY 215 -----GE-----CEVGVWIDE--GACVDYDECAAPR 238  
 DB 614 CPEPNIGNDCRHPCEALNGAHQAQMLANGQAQCLCAPYTGTSALAGCNDIDECRAN- 672  
 QY 239 PPGSAOQFCFNKNGSYTC-----EEC-DSSVGCTGEGP-----G 272  
 DB 673 -FCAEKAITSNTAGYLLCCPGGSSGDPYRBCGCTISKYTGCSANPCANGCTGVDSYTG 731  
 QY 273 N--CKECLISYAR--EHGQCADVDECSLAE-KTCVRRKNENCYTPGSIYCVCPDG 322  
 DB 732 NSVC-ICRGYERHNSENGCCODVDECSYGRKRAGLNLKLNLPSTSYECRCOG 785

RESULT 35  
 Q9NTM2 PRELIMINARY; PRT; 682 AA.  
 AC Q9NTM2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE BA124N14.3 (INTRINSIC FACTOR-B12 RECEPTOR) (FRAGMENT).  
 GN BA124N14.3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RL Bird C.;  
 RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.  
 CC -!- SIMILARITY: CONTRAINS 2 CUB DOMAINS.  
 DR EMBL: AL133415; CAB87965.1; -.  
 DR HSSP: P35555; IEMN.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR Pfam: PF00431; CUB; 2.  
 DR Pfam: PF00431; CUB; 2.  
 DR SMART: SM00042; CUB; 2.  
 DR SMART: SM00179; EGF\_CA; 3.  
 DR SMART: SM00001; EGF\_Like; 5.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 4.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_4.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_2.  
 DR PROSITE: PS01187; EGF\_CA; 3.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;  
 KW Receptor; Repeat.  
 FT NON\_TER 682  
 SQ SEQUENCE 682 AA; 74166 MW; 76D118A760E6A11 CRC64;

Query Match 12.6%; Score 252; DB 4; Length 682;  
 Best Local Similarity 29.1%; Pred No. 8.9e-16;  
 Matches 76; Conservative 26; Mismatches 95; Indels 64; Gaps 16;

QY 118 WPCVKTLLKVCSPGYGPDLA---CQ---GGSORPCSGNHCSDGSGRQDGSRCRCHNG 171  
 DB 154 FFCI-----CPQWKGPIGCSADVNECEIYSGPIGSCQNGTCV---NTMGSYSCHCPPE 204  
 QY 172 YQGPICLCT---DCMDGYFSSLRNETHSITACDESKTCSGLTNRCGCEVGVWL--DE 225  
 DB 205 TYGPQCAKRYDDCEGG--SVARCVHGI--CEDLMREGAGEPKYSC--VCDAGMWFSPNS 257  
 QY 226 GAC-VYVDCAAEAPPSCSAOFCFNKNGSYTCRECDSSCVGCTGEGPNCCKECLISYARE 284  
 DB 258 PACTLDRDECSFQPGFCSTLVOCFNTGSGFYGCACPTGNG-----NKYI-- 302  
 QY 285 HGOQADVDECSLAETKC-VKKNENCYTPGSIYVC-VCPDGF-----ETEDA 329  
 DB 303 ---CEDINCEIINNGGCSAAPVECVNTPGSSHCQACPGYQGDGRVCTLTIDICSVSNG 359  
 QY 330 CVPPAEAEATGESF--TOLP 348  
 DB 360 CHPDASCSTLGLSLPLCTCLP 380  
 RESULT 36  
 Q99K58 PRELIMINARY; PRT; 1174 AA.  
 AC Q99K58;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE SIMILAR TO FIBOLIN 2.



OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN NCBL\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD. GROSS  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC005443; AAH05443.1; -.  
 DR HSP: P00736; IAPQ.  
 DR Interpro: IPR000020; Anaphylatoxin.  
 DR Interpro: IPR000152; Asx\_hydroxyl.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR001881; EGF\_Ca.  
 DR Pfam: PF01821; ANATO\_2.  
 DR Pfam: PF00008; EGF\_6.  
 DR SMART: SM00104; ANATO\_3.  
 DR SMART: SM00181; EGF\_11.  
 DR SMART: SM00179; EGF\_CA\_9.  
 DR PROSITE: PS01177; ANAPHYLATOXIN\_1; 3.  
 DR PROSITE: PS01178; ANAPHYLATOXIN\_2; 3.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 5.  
 DR PROSITE: PS01186; EGF\_2; 5.  
 DR PROSITE: PS01187; EGF\_CA; 9.  
 KM Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SQ SEQUENCE 1174 AA; 126460 MW; 80628AC710FBA68 CRC64;

Query Match 12.5%; Score 251.5; DB 11; Length 1174;  
 Best Local Similarity 24.5%; Pred. No. 1.8e-15;  
 Matches 95; Conservative 29; Mismatches 123; Indels 141; Gaps 19;

QY 83 CESS-----DFECNQL-----EAGEHELEAMMLQLS 110  
 DB 525 CESPNLGYPCNHYMLSCCEGEPLIVEPARPEEAAPRVSETEMASREALSLGTEA 584  
 QY 111 EYP-----DLFEWCVKTL---KVCCSPGY-----GPDCLACGGSG-- 145  
 DB 585 ELPLNSLGGDDDECLMLPGEILCOHLCTIVGSTRACAFPEGLGDDGHTCPDKAPOLD 644  
 QY 146 -----RPSGNGHCSGDSRGD--GSCRCMHGY-----QG 174  
 DB 645 TAREADRESAQSPTIPLVPQPMTCXKNGPCRYCVRVGTAMSCSPGYAINADG 704  
 QY 175 PLCTD---CHDGFSSLRNETHSICACDECKTCGILTRNDCGECEYGVWLDGACVDV 231  
 DB 705 VSCDINECVTLHTCTRAE--HCVNTPGSFQC--YKALT-----CEPGYVLTGDECTDV 755  
 QY 232 DECAAEPPCSAAQFCCKNANGSYTCCECDSSCVGTGEGPNC---KECIS----- 279  
 DB 756 DECVTGHNCQAGFSQNTKGSFYCQARQRCMDGFLDDPRENCYDINECTSLLEPCRSGF 815  
 QY 280 -----GY-AREHG-OCADYDECSLAETKCVRKNECYNTPGSYVC 317  
 DB 816 SCINTVGSYTCQARNPLVCGRGYHANREGSECVVNECETGVHRC--GEGQLCYNLPGSYRC 874  
 QY 318 VCPGFEETEDA-----CVPPAEATG 341  
 DB 875 DCKPGFO--RDAFGRTCIDVNECVSPG 900

RESULT 37  
 008999  
 ID 008999 PRELIMINARY; PRT; 1833 AA.  
 AC 008999;  
 DT 01-JUL-1997 (TREMblrel. 04, Created)  
 DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE LATENT TGF-BETA BINDING PROTEIN-2.  
 GN LTBP2.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN NCBL\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RA Fang J., Li X., Smiley E., Francke U., Meacham R.P., Bonadio J.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF004874; AAB61611.1; -.  
 DR HSP: P35555; IEMN.  
 DR MGD: MGI:99502; Ltbp2.  
 DR Interpro: IPR002086; Aldehyde dehydr.  
 DR Interpro: IPR000152; Asx\_hydroxyl.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR001881; EGF\_Ca.  
 DR Interpro: IPR002212; TB.  
 DR Pfam: PF00008; EGF\_16.  
 DR Pfam: PF00683; TB; 4.  
 DR SMART: SM00179; EGF\_CA; 16.  
 DR SMART: SM00001; EGF\_like; 4.  
 DR PROSITE: PS00070; ALDEHYDE DEHYDR\_CYS; UNKNOWN\_1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 12.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; 10.  
 DR PROSITE: PS01187; EGF\_CA; 16.  
 KM Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SQ SEQUENCE 1833 AA; 197928 MW; 3FCD6BD31E8EFC6E CRC64;

Query Match 12.5%; Score 250; DB 11; Length 1833;  
 Best Local Similarity 24.8%; Pred. No. 4.3e-15;  
 Matches 82; Conservative 22; Mismatches 94; Indels 132; Gaps 16;

QY 128 CSPG-----TYGPDLACGGSGORPCSGNGHCSGDSRGDSCRCMHGY-----QG 174  
 DB 883 CSPGQLHPGODYCTDNECH---RNPEGRCV---NSVGSYCLCTPGITLYLTADT 936  
 QY 175 PLCTD---CHDGFSSLRNETHSIC-----TACDESC 203  
 DB 937 QECQDIDCEGEPGVCGSGRCNTGESSYCECDRGYIMVRKHQDINCRHPTGCPDRC 996  
 QY 204 KTCGSLNRDCGECEYGVWLDGACVDVDE-----CAEP----- 238  
 DB 997 VNSPG--SYTLACEEGYVGGSGCVDNECTPGICTHGRICINMEGSRCSCEPGYEV 1054  
 QY 239 -----PPCSA-----AFCNANGSYTCCEEDS-----SCVG----- 265  
 DB 1055 PDKKGRVDVDCASRASCPTGLCLNTGESSFYCSACQSGYWNEDGTACEDLDECAFPVC 1114  
 QY 266 ---CTGE--GPNGKECTISGYARE--HGQCADVDCECSLAETKCVRKNECYNTPGSYVC 318  
 DB 1115 PTGCTNIVGFSFSCDKDQGGYRPNPLGNRCEDVDCEBPOSSC--RGCEKNTGSSYCL 1172  
 QY 319 CPDGEET-----EDACVPPAE 335  
 DB 1173 CHOGFOLVNGTMCEDVNECVGEHCAHPGE 1202

RESULT 38  
 096R09  
 ID 096R09 PRELIMINARY; PRT; 3494 AA.  
 AC 096R09;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE INTRINSIC FACTOR-VITAMIN B12 RECEPTOR (FRAGMENT).  
 GN CUBN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN NCBL\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RA Aminoff M., Brady S., Verroust P.J., Moestrup S.K., Krahe R.;



OY	226	GAC -VDVDECAAEPPPCGAAPFCRKANSYTCBDDSSCVCTGGPBNCKECLISGYARE	284
Dd	258	PACTTDRRECSFQBPCCSTLVQCFTMGCSFYCGACPTMQG-----NGYT--	302
OY	285	HGCADVDECSLAETKC-VRKNENCYNPGSYVC-VCPDGEF-----ETIDA	329
Dd	303	---CSDINECEELNNCGSCVAPPVECVNTPGSSHOCACPRYGODORVCTLTDICSVSNG	359
OY	330	CVPAAEAATEGESF--TQLP	348
Dd	360	CHPDASCSTSLGSPLCTCLP	380
RESULT	40		
O9WUH9			
ID	O9WUH9	PRELIMINARY;	PRT; 2906 AA.
AC	O9WUH9;		
DT	01-NOV-1999 (TREMBLrel. 12, created)		
DT	01-NOV-1999 (TREMBLrel. 12, last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)		
DE	FIBRILITIN-2.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RP	[1]		
RA	SEQUENCE FROM N.A.		
RX	MEDLINE=99350231; PubMed=10419698;		
RA	Yang Q., Ota K., Tian Y., Kumar A., Wada J., Kashihara N., Wallner E.,		
RA	Kanwar Y.S.;		
RT	"Cloning of rat fibrillin-2 cDNA and its role in branching		
RL	morphogenesis of embryonic lung."		
DEV	Biol. 212:229-242(1999) .-		
EMBL	A0135060; AAD34439.1; .-		
HSSP	P35555; IEMN.		
DR	InterPro: IPRO02086; Aldehyde_dehydr.		
DR	InterPro: IPRO00152; Asx_hydroxyl.		
DR	InterPro: IPRO00561; Asx_hydroxyl.		
DR	InterPro: IPRO01881; EGF_CA.		
DR	InterPro: IPRO01438; EGF_II.		
DR	InterPro: IPRO02212; TB.		
DR	Pfam: PF00008; EGF; 46.		
DR	Pfam: PF00683; TB; 9.		
DR	PRINTS: PRO0010; EGFBLOOD.		
DR	SMART: SM00179; EGF_CA; 42.		
DR	SMART: SM00001; EGF_Like; 4.		
DR	PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.		
DR	PROSITE: PS00010; ASX_HYDROXYL; 43.		
DR	PROSITE: PS00022; EGF_1; UNKNOWN_2.		
DR	PROSITE: PS01186; EGF_2; 36.		
DR	PROSITE: PS01187; EGF_CA; 43.		
KW	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.		
SQ	SEQUENCE 2906 AA; 313371 MW; 9EE64E727044EF58 CRC64;		
Query Match	12.4%; Score 249; DB 11; length 2906;		
Best Local Similarity	22.0%; Pred. No. 9e-15;		
Matches	94; Conservative	54; Mismatches 121; Indels 158; Gaps	23;
OY	33	RCRGGLVDFR----NGM-VDTAKN-----FGGS--NT-AWEKTLKYE	69
Dd	1080	KCRNTIGSPKRCNSGFMALDMERNCTDIEDCRISPDLGNICVNTGSEFCEFEYE	1139
OY	70	SEIILLLELF-----GLCESSD--FEC-----NMLEADEFHEANWLQK	109
Dd	1140	SEFNMKKCMCIDECERNPDLRCRGTCVNTGESFOCDPLGHLSPRDCDIDIECSLS	1199
OY	110	SEYPDLFEWFC-----VKTLKVCSPGTG-PD-----CLMGQSGSRPSCSN	151
Dd	1200	DN-----LCNKGKCVNMNICTGYQCSNCPGYATFDRCGCSIDIDECAIMNGSCDTQCTNS	1252
OY	152	GHCSGDGSRHOGDGSCRHWGY-----GGPLCTD-----CM	181

[illegible]

```

Db 125 LMDDELNNNGGCOHICTNIVIGSYECRCKGFELSDNQHICIRSEGLS----- 174
Oy 113 POLFEMFECVTKLVG---CSPGTGYPDCLACOGG-----SQRPC-----SGNGHSGDGS 159
Db 175 -----CMKDKGCGCHICKKEAPRGVACBCECRGFEELANNQKCCILTCNHNNGGCCQHSCE 227
Oy 160 ROGDG-SCRCHMGVQ---GPLC-----TDCMDGYFSSLNETHSICT-- 197
Db 228 DTFEGEPCSCHPRYRLHADGRSCLDEQGYLVEGTESNATSVADGDKRYKRRLMERCAVN 287
Oy 198 ---ACDESKCTCSGLTNRDGCECEYGVNL--DEGACVYDVBCAAEPPECSAOCCKNANGS 253
Db 288 NGGCDRTCKDTS--TGVMHC-SCPTGFTLOYDGKTKDIDECOTRNGGCN--HFCKNTVGS 342
Oy 254 YTC-----EECDSSCVGCTGCGEPGNCCKEJCISGVARHGGCA 289
Db 343 FDCSCKKGKRLTDEKSCQDVBCSLERTCDHSCINHPTGTCACNPGTLYSFTH--CG 400
Oy 290 DVDCESLAEKTCVKNKNCNYPGYSVVCPPDGE--ETEDACVPPAEAEATEGESPTOL 347
Db 401 DINEGCVNNGGC---QOVCTNIVGSYECQCHPGFKLHMNKKDCV-----EVAGFPPTSM 451
Oy 348 PSREDL 353
Db 452 TPRVSL 457

```

## RESULT 42

```

Oy 09HB05 PRELIMINARY: PRT: 495 AA.
AC 09HB05;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 54.3 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBL_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel Human cDNA clones with function of inhibiting cancer cell
RT growth.";
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF217999; FAG17241.1; -.
DR HSSP: P35355; IEMN.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PR00008; EGF_6.
DR SMART: SM00181; EGF_9.
DR SMART: SM00179; EGF_CA; 9.
DR SMART: SM00001; EGF-like; 1.
DR PROSITE: PS00010; ASX_HYDROXYL.
KW EGF-like domain; glycoprotein; Hydroxylation; Hypothetical protein.
SQ SEQUENCE 495 AA; 54340 MW; C40434B6C82E3D70 CRC64;

```

Query Match 12.3%; Score 246.5; DB 4; Length 495;  
 Best Local Similarity 28.4%; Pred. No. 2, 1e-15;  
 Matches 83; Conservative 23; Mismatches 97; Indels 89; Gaps 17;

```

Oy 120 CVYTLVCC--SPGTGYPDCLACOGSGRP-----CGSGHCSGDGS 159
Db 6 CGGVFACCCCKKSOETGDLV---GGIQETDKLIEVEEODEPYLNDRCGGGPKCKQOCR 61
Oy 160 ROGDG-SCRCHMGVQ---GPLCTD---CMDGYFS-----SLNETHSIC--- 196
Db 62 DTGDEYVVCSCFVGTQLLSDVSCEDYNECTTGSCHLGECSCTNIVGSFRCQDSSCGTG 121

```

```

Oy 197 --TACDESK---TC-SGLTN-----RDGCEGCVGLDE-GACVYDVE 233
Db 122 YELTEDNSCKDIDECESGHNCLPDFICQNTLGSFRCRKLQCKSGFIQDALGNCIDNE 181
Oy 234 CAAEPPPCSAOFCCKNANGSYTCEECDDSSVCGTGGPGNCKEJCISGY-AREHG-QCADV 291
Db 182 CLTSAPCEPFGHCINTFESYTCOK-----NVPCGSGYHINEGTCVVDV 227
Oy 292 DECSLAEKTCVKNKNCNYPGYSVVCPPDGE--FEETEDACVPPAEATEG 341
Db 228 DECAPPAEPC-GKHRCVNSPSSFRCECKTGYTFDEISRMCAVDYNECQRRP 278

```

## RESULT 43

```

Oy 020911 PRELIMINARY: PRT: 3871 AA.
AC 020911; Q23242;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ZC116.3 PROTEIN.
GN ZC116.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBL_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Burton J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Faveille A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan R., Waterston R.,
RA Watson A., Weinstock K., Wilkinson-Spoat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 18 CUB DOMAINS.
DR EMBL: Z74473; CAA98952.1; -.
DR EMBL: Z74046; CAA98952.1; JOINED.
DR EMBL: Z74046; CAA98557.1; -.
DR EMBL: Z74473; CAA98557.1; JOINED.
DR HSSP: P00740; IEDM.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 13.
DR Pfam: PR00008; EGF_7.
DR SMART: SM00042; CUB; 18.
DR SMART: SM00179; EGF_CA; 3.
DR SMART: SM00001; EGF-like; 5.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 20.
DR PROSITE: PS00002; EF_HAND; UNKNOWN_4.
DR PROSITE: PS01186; EGF_2; 2.

```

DR PROSITE; PS01187; EGF\_CA; 3.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.  
 SO SEQUENCE 3871 AA; 433663 MW; 843600647421AB12 CRC64;

Query Match 12.2%; Score 245.5; DB 5; Length 3871;  
 Best Local Similarity 28.8%; Pred. No. 2,7e-14;  
 Matches 74; Conservative 26; Mismatches 94; Indels 63; Gaps 15;

OY 128 CSPGTGPDCLA-----CG--GSGRPGSGNGHCSGDSRGDSCGRHMKGQGFGLTCDM 181  
 DB 100 CPHFTGTTTEADIDECSSVYNGTTAGCQNGTCT--NNRGFGCCQCGSYHGSLCQ--- 153  
 OY 182 DGYFSLRNETHSICTA---CDESCKTCGSLTNHGE-----CEVGWVL-----DEGACVD 230  
 DB 154 --YHMSACSTFELCGPHGHCIESIYDPTGQSSDITTYKICIDMGKRVSSDKKNPTCYD 211  
 OY 231 VDECAEPSPCSAAQFCNNANGSYTCECDSS-----CVG---CTGE----- 269  
 DB 212 VNEC--ESNCPHGVDCINLPGSFVSCGPKGYKTDGNVCIDVNECGEIRVCSPLSKH 269  
 OY 270 ---GPGNCKCISGYAHEHQCADVDEC-----SLAEKTCVRKNENCYNTPGSYVCP 320  
 DB 270 NTLGSYYCDSCPTGYSGDGGNCVYKDCVKNKCHKLA--TC-KVTDGYSAGVDYTCYCP 326  
 OY 321 DGEFEET---EDACVPPA 334  
 DB 327 DGYVGDGIGEGCYKSA 343

## RESULT 44

025678 ID 025678 PRELIMINARY; PRT; 733 AA.

AC 025678; 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
 DE FIBRILLIN (FRAGMENT).  
 OS Podocoryne carnea.  
 OC Eukaryota; Metazoa; Chordata; Hydrozoa; Hydroida; Anthomedusae;  
 OC Hydroclitidae; Podocoryne.  
 OX NCBI\_TaxID=6096;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-95301106; PubMed-7781906;  
 RT Reber-Muller S., Spissinger T., Schuchert P., Spring J., Schmid V.;  
 RT "An extracellular matrix protein of jellyfish homologous to mammalian  
 RT fibrillins forms different fibrils depending on the life stage of the  
 RT animal";  
 RL Dev. Biol. 169:662-672(1995).  
 DR EMBL; L39930; AAA91336.1; -;  
 DR HSSP; P35555; 1EMN  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR001128; Cyt\_P450.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_CA.  
 DR InterPro: IPR001438; EGF\_II.  
 DR InterPro: IPR002212; TB.  
 DR InterPro: IPR001491; Thrombomodulin.  
 DR Pfam; PF00008; EGF; 14.  
 DR PRINTS; PR00010; EGFBL00D.  
 DR PRINTS; PR00907; THROMBOMODULN.  
 DR SMART; SM00017; EGF\_CA; 11.  
 DR SMART; SM00001; EGF\_like; 3.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 12.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 DR PROSITE; PS01186; EGF\_2; 11.  
 DR PROSITE; PS01187; EGF\_CA; 10.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 FT NON\_TER 1 1  
 FT NON\_TER 733 733  
 SO SEQUENCE 733 AA; 79856 MW; D65699052731D5E7 CRC64;

Query Match 12.2%; Score 244; DB 5; Length 733;  
 Best Local Similarity 25.1%; Pred. No. 5,7e-15;  
 Matches 92; Conservative 24; Mismatches 85; Indels 166; Gaps 21;

OY 111 EYPLPFMFVCKTLKVCSPGT---YGPCLAC-----OGSGRPGSGNGHCSGD--- 158  
 DB 171 EYPL-----CRHGSCSNKIGSFMCQNEGFKDQATNQSDINECQDPCK 218  
 OY 159 -----SRQDGSRCRCHNGYQ---GPICTD-----CMD 182  
 DB 219 NGNCRNRIGSAVCTCYGEYETIDGLSCDEKDECADENRQLGTCVNTDGFKLCNP 278  
 OY 183 GY-----FSSLRNE-----THSICTACD-----ESKTC--- 206  
 DB 279 GFVSDENERECIDVRSFCSSLENNOCTRANGLNVTKSYC-CCSMGAGFDDPELCPVK 337  
 OY 207 -----SGL-----TNRDGECEVG-----WVLDEG----- 226  
 DB 338 GTRFEYTCRNGVGFITSNNTDINECANLGTCLNGCQVNAVSFRVCDBGTYLNNP 397  
 OY 227 -ACVDVDECAAEPPPCSAOFCNNANGSYTCECDSSCVGCTGEGPCNCKECSIGYAREH 285  
 DB 398 RECIDRECAAGSNYCGNG-CTNLVGSYOC-SCEE-----GREPGT-----DS 439  
 OY 286 GQCADVDECSLAETCYRKNECNTPGSYVCPDGFEEEDA--CVPAEATGEGS 343  
 DB 440 PACVDVDECAKELCYRKNLV--GSYMSCPKGFKLAEDQKCV-----DINCDT 491  
 OY 344 PTOLPSR 350  
 DB 492 PHRCPR 498

## RESULT 45

09NO36 ID 09NO36 PRELIMINARY; PRT; 999 AA.

AC 09NO36; 01-OCT-2000 (T-EMBLrel. 15, Created)  
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
 DE CEGP1 PROTEIN.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bahr A., Hankeln T., Winterpacht A., Zabel B., Schmidt E.R.;  
 RT "Comparative sequencing of human chromosome 11p15 and mouse chromosome  
 RT 7.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
 DR EMBL; AJ400877; CAB92285.1; -;  
 DR HSSP; P07204; 1ADX.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_CA.  
 DR Pfam; PF00431; CUB; 1.  
 DR Pfam; PF00008; EGF; 5.  
 DR SMART; SM00042; CUB; 1.  
 DR SMART; SM00179; EGF\_CA; 6.  
 DR SMART; SM00001; EGF\_like; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 6.  
 DR PROSITE; PS01180; CUB; 1.  
 DR PROSITE; PS01186; EGF\_2; 8.  
 DR PROSITE; PS01187; EGF\_CA; 6.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SO SEQUENCE 999 AA; 109956 MW; 61334844A0053095 CRC64;

Query Match 12.2%; Score 244; DB 4; Length 999;  
 Best Local Similarity 25.1%; Pred. No. 8.1e-15;  
 Matches 120; Conservative 31; Mismatches 166; Indels 162; Gaps 28;

```

QY 7 AALGLPLLLLPAPAAKPPCH-RCRGL---VDKNGQVNPAPKNGGGNTAMEE 62
DB 11 AAMAVLLLLPLLLLAGAVPPGRRAAGPQEDVDECAOGL-DCHADALQONTPTSY 69
QY 63 KTLISK--YSSSEIRLEILEG-LCESSDFECQNMLEAQBHEHLEAMWLOKSEYP-DLFEEM 118
DB 70 KSCSKPGTGG-----EGROCEDID-ECGNELNGCVH---DCLNIPGNTICTCFDG 116
QY 119 F-----CYKTL---KYCCSPGYT-----GPD 136
DB 117 FMLAHGHNCLDDECLENNNGGQHTCVANVMGSECCCKEGFLSDNQHCHIRSEGLS 176
QY 137 CL-----ACO-----GGSQRPC-----SGNHCSDGSRQGDG-S 165
DB 177 CMNKHGCSHICKKADRGSAECRCRPFELAKNORCIIITCNHNGSCGCHSCDDTADGPE 236
QY 166 CRCHMGYO---GPLCTDCMDGYFSSLRNETHSI-----CT-----ACDS 202
DB 237 CSHHPYKMHHTDRSCLEBEDIVLEVTESNTTSYVDGDKRKRLMETCAVANNNGCDDRT 296
QY 203 CKTCSGLTRNDGCECEVGVW--LDEGACVVDDECAAEPPCSAAOFCCKNANGSYTC----- 256
DB 297 CKDTS--TGVMC-SCPVGFTLQDGKTKCDIDECQTRNGGCD--HFCKNIVSGFDGCKK 351
QY 257 -----EECDSSCVGCTGCGPNCCKEJCISGYAREHGQCADVDECSL 296
DB 352 GFKLITDEKSCODVDECSLDRCTDHSCINHPGFACACNNGYLYGPTH--CGDTRECSI 409
QY 297 AEKTCYRKNENCNTPEGSYVCVCPDGEF--ETPDACVPAEAATGESPTOLPSREDL 353
DB 410 NNGGC---QQVCYNTVGSIECGCHPCKLHMNKKDCY-----EVKGLPLTSPVPSVL 459

```

RESULT 46  
 091VK0 PRELIMINARY; PRT: 915 AA.

Query Match 12.1%; Score 243; DB 11; Length 915;  
 Best Local Similarity 29.5%; Pred. No. 9.1e-15;  
 Matches 65; Conservative 21; Mismatches 86; Indels 48; Gaps 13;

```

QY 125 KVC---CSPCTYGPDLACOGSGQRPCSGNGHSGDGSROGDSRCRCHMGY-----QGPL 176
DB 670 RIVSSCPREHYADKKRC-----RKCAPP--CESCFGSIGNCLCKTIGYFLNETSSC 722
QY 177 CTDCMDGYFSSLRNETHSICTACDESKCTCSGLTRNDGCECEVGVWLDDEGACVVDDECA 236
DB 723 VTCPCPGSYEDIK---NWCCKSEKCKACIGPHN--CTECKGGLSLQSGRC-----769
QY 237 EPPPCAAOFCCKNANGSYTCBECDDSCVGTGCGPNCCKEJCISGYAREHGQCADVDECSL 296

```

DB 770 -SVTCEDGQF---FNG-HDQCPCHRCATCGAGADGCINCTGYVMEBGRG--VQSCSV 822  
 QY 297 A-----EKTCYRKNENCY--NTPGSYVC-VCPODF 323  
 DB 823 SYLDHSSREGYKCKCKRCNDNSCLTCNGRPFKNCCSPSY 862

```

RESULT 47
QY 091V88 PRELIMINARY; PRT: 561 AA.
DB 01-DEC-2001 (TREMBLrel. 19, Created)
DB 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DB 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DB POEM (NEPHROECTIN SHORT ISOFORM).
DB GN POEM OR NEPH.
DB Mus musculus (Mouse).
DB Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
DB Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
DB NCBI_TaxId=10090;
DB [1]
DB SEQUENCE FROM N.A.
DB STRAIN=C57BL/6;
DB MEDLINE=21551216; PubMed=11546798;
DB Morimura N., Tezuka Y., Watanabe N., Yasuda M., Miyatani S.,
DB Hozumi N., Tezuka K.;
DB "Molecular cloning of POEM. A novel adhesion molecule that interacts
DB with alphabeta1 integrin".
DB J. Biol. Chem. 276:42172-42181(2001).
DB [2]
DB SEQUENCE FROM N.A.
DB STRAIN=NIH SWISS; TISSUE=KIDNEY;
DB MEDLINE=21363579; PubMed=11470831;
DB Bradenberger R., Schmidt A., Linton J., Wang D., Backus C., Denda S.,
DB Muller U., Reichardt L.F.;
DB "Identification and characterization of a novel extracellular matrix
DB protein nephronectin that is associated with integrin alphabeta1 in
DB the embryonic kidney."
DB J. Cell Biol. 154:447-458(2001).
DB EMBL: AB059656; BAB69692.1; -.
DB EMBL: AY035898; AAK96010.1; -.
DB SEQUENCE 561 AA; 61490 MW; 69E7ACAADEB3F506 CRC64;

```

Query Match 12.0%; Score 241.5; DB 11; Length 561;  
 Best Local Similarity 25.3%; Pred. No. 7.3e-15;  
 Matches 74; Conservative 25; Mismatches 113; Indels 81; Gaps 12;

```

QY 105 WLQKSEYDLEFWFCVTKLVCCSPGYGP--DCL-----ACQGGSGRPPSGNGH 153
DB 14 YLOVADFDGRMPROIYSSIGLC---RYGRIIDCWMGAROSGQCOFVCOPOCK-MGE 68
QY 154 CSGDGSROGDSRCRCHMGYQGLCTD-----CMDGYF-- 185
DB 69 C-----VPPNCKCKHPGAGKTCNODLNEGKLPKPPCKNHRQMTFGSKYKCLMGMLL 122
QY 186 --SSLNETHSICTACDESKCTCSGLTRNDGCECEVGVWLDGACVVDDECAAEPPPSA 243
DB 123 PDSGSSALSCSMANQCYCDVYKQCPSPGLQLPDRCTCVDIDECATGRVSCR 182
QY 244 ADFCKNANGSYTCBECDDSCVGTGCGPNCCKEJCISG---YAREHGQCADVDECSLAEK 299
DB 183 FRCQCVTFGSLYC-----KCHTGFPLMYIGKKYQCHDIDDECSLQGH 223
QY 300 TCVRKKNENYNTPGSYVCVCPDGEFETEDACV--PPAEATEGSGSPQLPSR 350
DB 224 QC--SSTARCYNIHGSYKCCQCRDGYEGDGLNCVYIPKVMIEPS---GPIHMPER 272

```

RESULT 48  
 090H16 PRELIMINARY; PRT: 554 AA.  
 ID 090H16  
 AC 090H16;





```
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE ENDOPROTEASE FURIN.
GN FURIN.
OS Spodoptera frugiperda (fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuidae; Noctuidae; Amphipyridae; Spodoptera.
OX NCBI_Taxid=7108;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SF9, FALL ARMYWORM OVARY;
RA Clepik M., Klenk H.;
RT "Cloning and functional characterization of FURIN from Spodoptera
RT frugiperda (SF9) cells.";
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z68888; CAA93116.1; -.
DR HSSP; Q99405; IMPT.
DR MEROPS; S08.0PB; -.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002029; Peptidase_S8.
DR InterPro; IPR002884; P_domain.
DR Pfam; PF01483; P; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTLISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM00261; FU; 10.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Protease.
KW SEQUENCE
SQ 1299 AA; 142020 MW; 4C3799C7BBC572AB CRC64;

Query Match 11.9%; Score 239; DB 5; Length 1299;
Best Local Similarity 25.8%; Pred. No. 3.3e-14;
Matches 92; Conservative 32; Mismatches 146; Indels 86; Gaps 20;

QY 19 PPAPEAKKPTPCRCGLVDKFNQGNVDTAKKNGGNTAWMEKTL--KYE-SSEIR 74
DB 881 PPAIYADKKRKECMRP-----VGCSTCTSAFCLSCPEKWLKKK 922
QY 75 LLEILGLCESSDF-----ECNOMLEA---QEEHLEAMWLQKSEYDPLF-EMFCYK 122
DB 923 CMFVGSDDKCSAGEFAVDQCKRCNPACDSCYGENEGHC-----LTCPNPNILODYKCV 976
QY 123 TLKVCSPGTGYP--DCLACQGGSGRPGCSGNHCSGDGSGRQDSCGRCHMGYQPLC-ID 179
DB 977 E---CSKGYAEAGRCARCMHG-----CSDCVSRINCTSCASTLRLOSGACRTS 1022
QY 180 CMDGYFSSLNETHSICTACDESKCTCSGLTNRDCGECEYGVILDEGACVDVDECAEAPP 239
DB 1023 CADGYTAD-----RGTCSKCYLSCRTCTGPRRQOCASCPBGWRLAAGEC-----HP 1068
QY 240 PCSAAQFCNNANGSYCEECSSCVCTEGPNCKECISGYAREHGQCAD-----VDE 293
DB 1069 ECPQG--FYQSPG---CRHCHHCRCDGSGPLHCKSCPFRMLDGLCMEICGSGQYDA 1124
QY 294 CSLAEKTCYAKNKNCTNFGSYVC--VC--PDGFEETEDACVP--PARAEATGESPT 345
DB 1125 TSGTCRSCDASCRTG--SGRQGFSTCTCSRPLRLIDRLNOCVPCCSERGVNTSTPPT 1179
```

Search completed: September 10, 2002, 11:13:37  
Job time: 254 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 11:09:23 : Search time 13.41 seconds  
(without alignments)  
1019.239 Million cell updates/sec

Title: US-09-905-075-2

Perfect score: 2005

Sequence: 1 MRLPRRAIGLEPLILLPP.....AEATEGEGSPQLPSREDL 353

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SwissProt\_40.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	Match Length	ID	Description
1	286.5	14.3	684	1 FBL1_CHICK	073775 gallus gall
2	274.5	13.7	2871	1 FBN1_PIG	09tv36 sus. scrofa
3	270	13.5	2871	1 FBN1_BOVIN	P98133 bos. taurus
4	268.5	13.4	2871	1 FBN1_HUMAN	P35555 homo. sapien
5	263	13.1	1877	1 PCK5_MOUSE	004592 mus. musculu
6	254.5	12.7	703	1 FBL1_HUMAN	P23142 homo. sapien
7	253.5	12.6	705	1 FBL1_MOUSE	008819 mus. musculu
8	252.5	12.6	2871	1 FBN1_MOUSE	061554 mus. musculu
9	252.5	12.6	2907	1 FBN2_MOUSE	061555 mus. musculu
10	249.5	12.4	2911	1 FBN2_HUMAN	P35556 homo. sapien
11	248.5	12.4	1696	1 PCK5_BRACL	09n115 brachiolesto
12	247	12.3	675	1 PRPS_RAT	P53813 rattus norv
13	247	12.3	1221	1 FBL2_MOUSE	P37889 mus. musculu
14	242.5	12.1	1184	1 FBL2_HUMAN	P98035 homo. sapien
15	240.5	12.0	1394	1 TGFBR_HUMAN	P22064 homo. sapien
16	239	11.9	676	1 PRPS_HUMAN	P07225 homo. sapien
17	239	11.9	1877	1 PCK5_RAT	P41413 rattus norv
18	236	11.8	913	1 PCK5_HUMAN	092814 rattus norv
19	232	11.6	675	1 PRPS_MOUSE	008761 mus. musculu
20	232	11.6	1375	1 NID2_MOUSE	014112 homo. sapien
21	228.5	11.4	649	1 PRPS_MACMU	028520 macaca mula
22	228	11.4	1680	1 FUR2_DROME	P30432 drosophila
23	226	11.3	2531	1 NTC1_MOUSE	001705 mus. musculu
24	225.5	11.2	646	1 PRPS_RABIT	P98118 oryctolagus
25	224.5	11.2	1403	1 NID2_MOUSE	088322 mus. musculu
26	223	11.1	2524	1 NTC1_XENLA	P21783 xenopus lae
27	222.5	11.1	886	1 EMRI_HUMAN	014246 bos. sapien
28	222	11.1	675	1 PRPS_BOVIN	P07224 bos. taurus
29	222	11.1	1712	1 TGFBR_RAT	000918 rattus norv
30	222	11.1	2444	1 NTC1_HUMAN	P46511 homo. sapien
31	221	11.0	1247	1 NID2_HUMAN	P14553 homo. sapien
32	220	11.0	2531	1 NTC1_RAT	007008 rattus norv
33	217.5	10.8	3672	1 LMT2_CAEEL	021313 caenorhabdl

## ALIGNMENTS

RESULT 1	ID	FBLL_CHICK	STANDARD:	PRT:	684 AA.
AC	073775;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Fibulin-1 precursor.				
GN	FBLN1.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OX	Gallus.				
RN	NCBI_TaxID=9031.				
RP	SEQUENCE FROM N.A.				
RA	Barth J.L., Argaves K.M., Roark E.F., Little C.D., Argaves W.S.;				
RT	Characterization of chicken and nematode fibulin-1 homologs and				
RT	functionalization of the nematode fibulin-1 gene."				
RL	Submitted (FEH-1998) to the EMBL/GenBank/DBJ databases.				
CC	- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.				
CC	- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.				
CC	- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
DR	EMBL: AF051400; AAC05388.1; -				
DR	HSSP: P00742; IHCG.				
DR	InterPro: IPR000020; Anaphylatoxin.				
DR	InterPro: IPR000152; Asx_hydroxyl.				

34	216.5	10.8	2703	1 NTC1_DROME	P07207 drosophila
35	216	10.8	448	1 FBL5_RAT	09wv18 rattus norv
36	214.5	10.7	448	1 FBL5_MOUSE	09wv19 mus. musculu
37	213.5	10.6	1245	1 NID2_MOUSE	P10493 mus. musculu
38	212	10.6	2318	1 NTC3_MOUSE	061962 mus. musculu
39	211.5	10.5	810	1 NTC1_HUMAN	092832 homo. sapien
40	211.5	10.5	931	1 EMRI_MOUSE	061549 mus. musculu
41	211.5	10.5	2437	1 NTC1_BRARE	P46530 brachydanio
42	209.5	10.4	448	1 FBL5_HUMAN	09wv15 homo. sapien
43	209.5	10.4	687	1 VS41_GIALA	P92127 giardia lam
44	209	10.4	2139	1 CRB_DROME	P10040 drosophila
45	208	10.4	443	1 FBL4_CRGR	055058 cricetus
46	207	10.3	443	1 FBL4_MOUSE	05wv19 mus. musculu
47	205.5	10.2	473	1 FP2_MYTGA	025464 mytilus gall
48	204.5	10.2	810	1 NTC1_RAT	062919 rattus norv
49	204.5	10.2	1964	1 NTC1_MOUSE	P31695 mus. musculu
50	203	10.1	443	1 FBL4_HUMAN	095967 homo. sapien
51	202.5	10.1	356	1 TRBM_BOVIN	P06579 bos. taurus
52	202.5	10.1	712	1 FBL1_CAEEL	077469 caenorhabdl
53	202.5	10.1	816	1 NTC1_RAT	062918 rattus norv
54	202	10.1	667	1 TS11_GIALA	003185 giardia lam
55	201.5	10.0	816	1 NTC1_MOUSE	061220 mus. musculu
56	201	10.0	353	1 DLK_HUMAN	P80370 homo. sapien
57	201	10.0	433	1 FBL3_HUMAN	012805 homo. sapien
58	201	10.0	835	1 CD97_HUMAN	P48960 homo. sapien
59	201	10.0	4289	1 TENX_HUMAN	P22105 homo. sapien
60	200.5	10.0	1064	1 FBL1_STRPU	P10079 strongylocet
61	200.5	10.0	4655	1 LRP2_HUMAN	P98164 homo. sapien
62	200	10.0	713	1 TS44_GIALA	P21849 giardia lam
63	197	9.8	577	1 TRBM_MOUSE	P15306 mus. musculu
64	197	9.8	723	1 DL11_HUMAN	000548 homo. sapien
65	196.5	9.8	816	1 NTC1_HUMAN	099435 homo. sapien

DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF-Ca.  
 DR Pfam; PF01821; ANATO; 2.  
 DR Pfam; PF00008; EGF; 6.  
 DR SMART; SM00104; ANATO; 3.  
 DR SMART; SM00179; EGF-Ca; 8.  
 DR SMART; SM00001; EGF-like; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 5.  
 DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
 DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 2.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 3.  
 DR PROSITE; PS01187; EGF-Ca; 8.  
 DR Signal; Glycoprotein; Extracellular matrix; Repeat; EGF-like domain;  
 KM Calcium-binding.  
 FT CHAIN 1 25 POTENTIAL.  
 FT SIGNAL 1 25 FIBRIN-1.  
 FT DOMAIN 33 74 ANAPHYLATOXIN-LIKE 1.  
 FT DOMAIN 75 109 ANAPHYLATOXIN-LIKE 2.  
 FT DOMAIN 110 142 ANAPHYLATOXIN-LIKE 3.  
 FT DOMAIN 177 216 EGF-LIKE 1.  
 FT DOMAIN 217 262 EGF-LIKE 2.  
 FT DOMAIN 263 308 EGF-LIKE 3.  
 FT DOMAIN 309 356 EGF-LIKE 4.  
 FT DOMAIN 357 399 EGF-LIKE 5.  
 FT DOMAIN 400 441 EGF-LIKE 6.  
 FT DOMAIN 442 481 EGF-LIKE 7.  
 FT DOMAIN 482 525 EGF-LIKE 8.  
 FT DOMAIN 526 570 EGF-LIKE 9.  
 FT DISULFID 33 59 BY SIMILARITY.  
 FT DISULFID 34 66 BY SIMILARITY.  
 FT DISULFID 47 67 BY SIMILARITY.  
 FT DISULFID 76 107 BY SIMILARITY.  
 FT DISULFID 89 108 BY SIMILARITY.  
 FT DISULFID 110 134 BY SIMILARITY.  
 FT DISULFID 111 141 BY SIMILARITY.  
 FT DISULFID 124 142 BY SIMILARITY.  
 FT DISULFID 181 191 BY SIMILARITY.  
 FT DISULFID 187 200 BY SIMILARITY.  
 FT DISULFID 202 215 BY SIMILARITY.  
 FT DISULFID 221 234 BY SIMILARITY.  
 FT DISULFID 228 243 BY SIMILARITY.  
 FT DISULFID 249 261 BY SIMILARITY.  
 FT DISULFID 267 280 BY SIMILARITY.  
 FT DISULFID 274 289 BY SIMILARITY.  
 FT DISULFID 295 307 BY SIMILARITY.  
 FT DISULFID 313 326 BY SIMILARITY.  
 FT DISULFID 320 335 BY SIMILARITY.  
 FT DISULFID 342 355 BY SIMILARITY.  
 FT DISULFID 361 374 BY SIMILARITY.  
 FT DISULFID 368 383 BY SIMILARITY.  
 FT DISULFID 385 398 BY SIMILARITY.  
 FT DISULFID 404 416 BY SIMILARITY.  
 FT DISULFID 412 425 BY SIMILARITY.  
 FT DISULFID 427 440 BY SIMILARITY.  
 FT DISULFID 446 455 BY SIMILARITY.  
 FT DISULFID 451 464 BY SIMILARITY.  
 FT DISULFID 466 480 BY SIMILARITY.  
 FT DISULFID 486 499 BY SIMILARITY.  
 FT DISULFID 495 508 BY SIMILARITY.  
 FT DISULFID 510 524 BY SIMILARITY.  
 FT DISULFID 530 543 BY SIMILARITY.  
 FT DISULFID 537 552 BY SIMILARITY.  
 FT DISULFID 557 569 BY SIMILARITY.  
 FT CARBOHYD 96 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 638 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 684 AA; 75623 MW; 1638D7A094739199 CRC64;

Query Match 14.3%; Score 286.5; DB 1; Length 684;  
 Best Local Similarity 25.7%; Pred. No. 2.8e-14;

Matches 127; Conservative 37; Mismatches 120; Indels 211; Gaps 29;  
 QY 6 RAALGLPLLLPLPPAPAAKKPFRCHRCGLVDKFNQGMVDTAKKNGGNTAMEEKTLL 65  
 DB 5 RQARPLRLILLILLALLPALRGODLSMECC---CDK---GYEMANKN----- 44  
 QY 66 SKYESSEIRLEILEGLCESSDFECNQMLEAO-----EHLNEMWILQKSEYDPL---- 115  
 DB 45 -----RICTSPLISESR--ECS-WTQYCCRSKLEHYCSGIDIFASVHECDSDHN 93  
 QY 116 -----FEMF-----CYKTLKV-----CCSGTGPDDL- 138  
 DB 94 GENSTCEAEYFKCCCYCLLKTAKVAGQSCPEPNKIKIYQCGIYFRACCVAGGQGTVDSI 153  
 QY 139 -----ACGGSGQRPCSGNGHCSGDSRQGDGSCRHMGY 172  
 DB 154 SDDAKRKQVEISKRELDQEDPYLHDCGRGG--PCS--QQCRDGGSY---VCSDFYGY 206  
 QY 173 Q---GPLCTD---CMDGYFSLRNETHS--ICTAC-----DESKTCSGLTN- 211  
 DB 207 QLPDGVNCCEDINECITG-----THSCGIGTCVNTLGSFRCQRTSGTGYELTDD 258  
 QY 212 ---RCCGCEV-----GWYDE-GACVYVDECAAPR 239  
 DB 259 SNCKDIDCECTGTHNCPDPFICQNTPGSFRCRPRKLCQNMGEFIQALGNCIDINECLSTNM 318  
 QY 240 PCSAOPFCNANGSYTC-----EEDSSCGCTGE-----G 270  
 DB 319 PCPAQICINDGSTTCQRISSPCSGRGYHNEDETRCYVDYDECSSDPC-GEHVYICNG 377  
 QY 271 PGNCX-ECISGYARE--HGQCADVDEC-SLAETCYVRNENCCYNTPGSYVGVCPDGEFET 326  
 DB 378 PGNRYCECKSGSFYDVISRTCIDINECHRRYPRGLCAHR---CENTPGSYCTCTGWFYLS 434  
 QY 327 ED--ACVPAFAEAT 339  
 DB 435 SDGRSCEDLNECESS 449  
 RESULT 2  
 FBN1\_PIG STANDARD; PRT; 2871 AA.  
 AC Q9TV36;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Fibrillin 1 precursor.  
 GN FBN1.  
 OS Sus scrofa (pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=99156858; PubMed=10036187;  
 RA Biery N.J., Eldadah Z.A., Moore C.S., Stetten G., Spencer F.,  
 RA Dietz H.C.;  
 RT "Revised genomic organization of FBN1 and significance for regulated  
 gene expression.";  
 RL Genomics 56:70-77(1999).  
 CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS  
 CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE  
 CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.  
 CC -!- PFM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER  
 CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE  
 CC MICROFIBRILS.  
 CC -!- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING  
 CC EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-



FT	DOMAIN	1113	1154	EGF-LIKE 17, CALCIUM-BINDING.
CE	DOMAIN	1155	1106	EGF-LIKE 18, CALCIUM-BINDING.

```

CC
DR EMBL; L28748; AAA74122.1; -.
DR HSSP; P35555; IAPJ.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR002212; TB.
DR Pfam; PF00008; EGF; 46.
DR Pfam; PF00683; TB; 9.
DR PRINTS; PRO0010; EGFBL00D.
DR SMART; SM00179; EGF_CA; 42.
DR PROSITE; PS00010; ASX_HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 38.
DR PROSITE; PS01187; EGF_CA; 45.
DR Extracellular matrix; GlycoCalc; 45.
KW Repeat; Signal; Multigene family.
KW SIGNAL.
FT CHAIN 1 27
FT CHAIN 28 2871
FT DOMAIN 81 112 FIBRILLIN 1.
FT DOMAIN 115 146 EGF-LIKE 1, NON-CALCIUM BINDING.
FT DOMAIN 147 178 EGF-LIKE 2, NON-CALCIUM BINDING.
FT DOMAIN 246 287 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 288 329 EGF-LIKE 5, CALCIUM-BINDING.
FT REPEAT 330 390 TGFBR 1.
FT DOMAIN 392 446 PRO-RICH.
FT DOMAIN 449 489 EGF-LIKE 6, NON-CALCIUM BINDING.
FT DOMAIN 490 529 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 530 571 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 572 612 EGF-LIKE 9, CALCIUM-BINDING.
FT DOMAIN 613 653 EGF-LIKE 10, CALCIUM-BINDING.
FT REPEAT 654 722 TGFBR 2.
FT DOMAIN 723 764 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 765 806 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 807 846 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 910 951 EGF-LIKE 14, CALCIUM-BINDING.
FT REPEAT 952 1027 TGFBR 3.
FT DOMAIN 1028 1069 EGF-LIKE 15, CALCIUM-BINDING.
FT DOMAIN 1070 1112 EGF-LIKE 16, CALCIUM-BINDING.
FT DOMAIN 1113 1154 EGF-LIKE 17, CALCIUM-BINDING.
FT DOMAIN 1155 1196 EGF-LIKE 18, CALCIUM-BINDING.

```

FT	DOMAIN	1197	1237	EGF-LIKE 19, CALCIUM-BINDING.
FT	DOMAIN	1238	1279	EGF-LIKE 20, CALCIUM-BINDING.
FT	DOMAIN	1280	1321	EGF-LIKE 21, CALCIUM-BINDING.
FT	DOMAIN	1322	1362	EGF-LIKE 22, CALCIUM-BINDING.
FT	DOMAIN	1363	1403	EGF-LIKE 23, CALCIUM-BINDING.
FT	DOMAIN	1404	1445	EGF-LIKE 24, CALCIUM-BINDING.
FT	DOMAIN	1446	1486	EGF-LIKE 25, CALCIUM-BINDING.
FT	DOMAIN	1487	1527	EGF-LIKE 26, CALCIUM-BINDING.
FT	REPEAT	1528	1605	TGFBP 4.
FT	DOMAIN	1606	1647	EGF-LIKE 27, CALCIUM-BINDING.
FT	DOMAIN	1648	1688	EGF-LIKE 28, CALCIUM-BINDING.
FT	REPEAT	1689	1765	TGFBP 5.
FT	DOMAIN	1766	1807	EGF-LIKE 29, CALCIUM-BINDING.
FT	DOMAIN	1808	1848	EGF-LIKE 30, CALCIUM-BINDING.
FT	DOMAIN	1849	1890	EGF-LIKE 31, CALCIUM-BINDING.
FT	DOMAIN	1891	1929	EGF-LIKE 32, CALCIUM-BINDING.
FT	DOMAIN	1930	1972	EGF-LIKE 33, CALCIUM-BINDING.
FT	DOMAIN	1973	2012	EGF-LIKE 34, CALCIUM-BINDING.
FT	DOMAIN	2013	2054	EGF-LIKE 35, CALCIUM-BINDING.
FT	REPEAT	2055	2126	TGFBP 6.
FT	DOMAIN	2127	2165	EGF-LIKE 36, CALCIUM-BINDING.
FT	DOMAIN	2166	2205	EGF-LIKE 37, CALCIUM-BINDING.
FT	DOMAIN	2206	2246	EGF-LIKE 38, CALCIUM-BINDING.
FT	DOMAIN	2247	2290	EGF-LIKE 39, CALCIUM-BINDING.
FT	REPEAT	2291	2332	EGF-LIKE 40, CALCIUM-BINDING.
FT	REPEAT	2333	2401	TGFBP 7.
FT	DOMAIN	2402	2443	EGF-LIKE 41, CALCIUM-BINDING.
FT	DOMAIN	2444	2484	EGF-LIKE 42, CALCIUM-BINDING.
FT	DOMAIN	2485	2523	EGF-LIKE 43, CALCIUM-BINDING.
FT	DOMAIN	2524	2566	EGF-LIKE 44, CALCIUM-BINDING.
FT	DOMAIN	2567	2606	EGF-LIKE 45, CALCIUM-BINDING.
FT	DOMAIN	2607	2647	EGF-LIKE 46, CALCIUM-BINDING.
FT	DOMAIN	2648	2687	EGF-LIKE 47, CALCIUM-BINDING.
FT	DISULFID	85	94	BY SIMILARITY.
FT	DISULFID	89	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	129	BY SIMILARITY.
FT	DISULFID	123	134	BY SIMILARITY.
FT	DISULFID	136	145	BY SIMILARITY.
FT	DISULFID	150	160	BY SIMILARITY.
FT	DISULFID	154	166	BY SIMILARITY.
FT	DISULFID	168	177	BY SIMILARITY.
FT	DISULFID	250	262	BY SIMILARITY.
FT	DISULFID	257	271	BY SIMILARITY.
FT	DISULFID	273	286	BY SIMILARITY.
FT	DISULFID	292	304	BY SIMILARITY.
FT	DISULFID	299	313	BY SIMILARITY.
FT	DISULFID	315	328	BY SIMILARITY.
FT	DISULFID	453	465	BY SIMILARITY.
FT	DISULFID	460	474	BY SIMILARITY.
FT	DISULFID	476	488	BY SIMILARITY.
FT	DISULFID	494	504	BY SIMILARITY.
FT	DISULFID	499	513	BY SIMILARITY.
FT	DISULFID	515	528	BY SIMILARITY.
FT	DISULFID	534	546	BY SIMILARITY.
FT	DISULFID	541	555	BY SIMILARITY.
FT	DISULFID	557	570	BY SIMILARITY.
FT	DISULFID	576	587	BY SIMILARITY.
FT	DISULFID	582	596	BY SIMILARITY.
FT	DISULFID	598	611	BY SIMILARITY.
FT	DISULFID	617	628	BY SIMILARITY.
FT	DISULFID	623	637	BY SIMILARITY.
FT	DISULFID	639	652	BY SIMILARITY.
FT	DISULFID	727	739	BY SIMILARITY.
FT	DISULFID	734	748	BY SIMILARITY.
FT	DISULFID	750	763	BY SIMILARITY.
FT	DISULFID	769	781	BY SIMILARITY.
FT	DISULFID	776	790	BY SIMILARITY.
FT	DISULFID	792	805	BY SIMILARITY.
FT	DISULFID	811	821	BY SIMILARITY.
FT	DISULFID	816	830	BY SIMILARITY.
FT	DISULFID	832	845	BY SIMILARITY.
FT	DISULFID	914	926	BY SIMILARITY.

FT	DISULFID	921	935	BY SIMILARITY.
FT	DISULFID	937	950	BY SIMILARITY.
FT	DISULFID	1032	1044	BY SIMILARITY.
FT	DISULFID	1039	1053	BY SIMILARITY.
FT	DISULFID	1055	1068	BY SIMILARITY.
FT	DISULFID	1074	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1111	BY SIMILARITY.
FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1124	1138	BY SIMILARITY.
FT	DISULFID	1140	1153	BY SIMILARITY.
FT	DISULFID	1159	1171	BY SIMILARITY.
FT	DISULFID	1166	1180	BY SIMILARITY.
FT	DISULFID	1182	1195	BY SIMILARITY.
FT	DISULFID	1201	1212	BY SIMILARITY.
FT	DISULFID	1208	1221	BY SIMILARITY.
FT	DISULFID	1223	1236	BY SIMILARITY.
FT	DISULFID	1242	1254	BY SIMILARITY.
FT	DISULFID	1249	1263	BY SIMILARITY.
FT	DISULFID	1265	1278	BY SIMILARITY.
FT	DISULFID	1284	1296	BY SIMILARITY.
FT	DISULFID	1291	1305	BY SIMILARITY.
FT	DISULFID	1307	1320	BY SIMILARITY.
FT	DISULFID	1326	1339	BY SIMILARITY.
FT	DISULFID	1333	1348	BY SIMILARITY.
FT	DISULFID	1350	1361	BY SIMILARITY.
FT	DISULFID	1367	1380	BY SIMILARITY.
FT	DISULFID	1374	1389	BY SIMILARITY.
FT	DISULFID	1391	1402	BY SIMILARITY.
FT	DISULFID	1408	1420	BY SIMILARITY.
FT	DISULFID	1415	1429	BY SIMILARITY.
FT	DISULFID	1431	1444	BY SIMILARITY.
FT	DISULFID	1450	1461	BY SIMILARITY.
FT	DISULFID	1456	1470	BY SIMILARITY.
FT	DISULFID	1472	1485	BY SIMILARITY.
FT	DISULFID	1491	1502	BY SIMILARITY.
FT	DISULFID	1497	1511	BY SIMILARITY.
FT	DISULFID	1513	1526	BY SIMILARITY.
FT	DISULFID	1610	1622	BY SIMILARITY.

Query Match 13.58; Score 270; DB 1; Length 2871;  
Best Local Similarity 24.88; Pred. No. 1.8e-12;  
Matches 107; Conservative 33; Mismatches 128; Indels 164; Gaps 25;

QY	19	PPAEBAKKRPKPP-----HRCGLVDR-----NCGMDTAKKNGGGGTA	59
DB	1141	PPGQPLAPNISACIDINECELSAHLCPHRCVNLIGKYOCACNPGYHSTPDLRF-----	1194
QY	60	WEKTLKSESESEIRLLEILEGLCES-----SDPECN-----OMLEAQEHLLEAMWLQ	108
DB	1195	-----CVDIDECISIMNGGCTPCTNSEGSEYSCQGFALMPQR-----	1234
QY	109	KSEYPDLEFMEFCVTKLYC-----CS--PGTYGPDLACOG-----GSORPCSGNGHC	154
DB	1235	--SCDIDDE--CEDPNICDGGQCTNIPGEY--RCLCYDGFMASEDMKTCVDVNECDLNP	1288
QY	155	-----SGD--GSNOGDSCHNGYOGPL-----CTDCMDGYFSSLRNETHSITACDESKT	205
DB	1289	NICLSGTENKRGSFICHDMDGYSGKRGKTCCTIDNECEIAGHNCDRHNAVCTMTAGSPK	1347
QY	206	CSGLTNRDCGCEGVWVDEGACVDVDECALEPPPCSAOFCKNANGSYTC-----	256
DB	1348	CS-----CSRGWIDGDKICTDLDSCSNGTHMCSOHAQCKNTGMSYRLCKRGTYGD	1398
QY	257	-----EEC-----DSSCVGCTG-----EGP-----	271
DB	1399	GFTCTDDECESENILNLCNGQCLNAPGGYRCECDMGFPVPSADGKCEDIDECSLPNICVF	1458
QY	272	GNCK-----ECISGA--REHQCADVDECSLAETCYRKKNENCNTPGSYVAVCPD	321
DB	1459	GTCHNLPGLFRCCEIGEILDRSGNGCTDVNEC--LDPTTCT--SGNCVNTGSGYSTCDDP	1515
QY	322	GFE--ETEDACV	331

DB 1516 DFEINPTFVGCV 1527

RESULT 4

FBNI\_HUMAN STANDARD; PRT; 2871 AA.

AC P35355;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Fibrillin 1 precursor.

GN FBNI OR FBNI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

[1]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RA MEDLINE=93372860; PubMed=8364578;

RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,

RA Pangillan T., Bonadio J.;

RT "Genomic organization of the sequence coding for fibrillin, the

RT defective gene product in Marfan syndrome.";

RL Hum. Mol. Genet. 2:961-968(1993).

[2]

RP SEQUENCE OF 1-932 FROM N.A.

RC TISSUE=Placenta, and Fibroblast;

RA MEDLINE=94010947; PubMed=7691719;

RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;

RT "Fibrillin binds calcium and is coded by cDNAs that reveal a

RT multidomain structure and alternatively spliced exons at the 5'

RT end.";

RL Genomics 17:476-484(1993).

[3]

RP SEQUENCE OF 899-2871 FROM N.A.

RA MEDLINE=91304568; PubMed=1852207;

RA Maslen C.L., Corson G.M., Maddox B.K., Glatville R.W., Sakai L.Y.;

RT "Partial sequence of a candidate gene for the Marfan syndrome.";

RL Nature 352:334-337(1991).

[4]

RP SEQUENCE OF 813-1313 FROM N.A.

RA MEDLINE=91304567; PubMed=1852206;

RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,

RA Tsipouras P., Ramirez F., Hollister D.W.;

RT "Linkage of Marfan syndrome and a phenotypically related disorder to

RT two different fibrillin genes.";

RL Nature 352:330-334(1991).

[5]

RP CHARACTERIZATION.

RA MEDLINE=91317849; PubMed=1860873;

RA Sakai L.Y., Keene D.R., Glatville R.W., Bachinger H.P.;

RT "Purification and partial characterization of fibrillin, a cysteine-

RT rich structural component of connective tissue microfibrils.";

RL J. Biol. Chem. 266:14763-14770(1991).

[6]

RP STRUCTURE BY NMR OF 2054-2125.

RA MEDLINE=98031893; PubMed=9362480;

RA Yuan X., Downing A.K., Knott V., Handford P.A.;

RT "Solution structure of the transforming growth factor beta-binding

RT protein-like module, a domain associated with matrix fibrils.";

RL EMBO J. 16:659-666(1997).

[7]

RP STRUCTURE BY NMR OF 2124-2205.

RA MEDLINE=96144829; PubMed=8568869;

RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;

RT "Calcium binding properties of an epidermal growth factor-like domain

RT pair from human fibrillin-1.";

RL J. Mol. Biol. 255:22-27(1996).

[8]

RP STRUCTURE BY NMR OF 2124-2205.

RA MEDLINE=96222301; PubMed=8653794;

RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,

RA Handford P.A.;

RT "Solution structure of a pair of calcium-binding epidermal growth

RT factor-like domains: implications for the Marfan syndrome and other

RT genetic disorders.";

RL Cell 85:597-605(1996).

[9]

RP REVIEW ON MFS VARIANTS.

RA MEDLINE=96174615; PubMed=8594563;

RA Colliod G., Beroud C., Soussi T., Junten C., Boileau C.;

RT "Software and database for the analysis of mutations in the human

RT FBNI gene.";

RL Nucleic Acids Res. 24:137-141(1996).

[10]

RP REVIEW ON MFS VARIANTS.

RA MEDLINE=97169383; PubMed=9016526;

RA Colliod-Beroud G., Beroud C., Agés L., Black C., Boxer M., Brock D.J.,

RA Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L.,

RA Richards R.I., Wang W., Junten C., Boileau C.;

RT "Marfan Database (second edition): software and database for the

RT analysis of mutations in the human FBNI gene.";

RL Nucleic Acids Res. 25:147-150(1997).

[11]

RP REVIEW ON VARIANTS.

RA MEDLINE=98062175; PubMed=9401003;

RA Hayward C., Brock D.J.H.;

RT "Fibrillin-1 mutations in Marfan syndrome and other type-1

RT fibrillinopathies.";

RL Hum. Mutat. 10:415-423(1997).

[12]

RP VARIANT MFS PRO-1137.

RA MEDLINE=91304569; PubMed=1852208;

RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,

RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,

RA Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;

RT "Marfan syndrome caused by a recurrent de novo missense mutation in

RT the fibrillin gene.";

RL Nature 352:337-339(1991).

[13]

RP VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.

RA MEDLINE=93250834; PubMed=1301946;

RA Dietz H.C., Saraiya J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;

RT "Clustering of fibrillin (FBNI) missense mutations in Marfan syndrome

RT patients at cysteine residues in EGF-like domains.";

RL Hum. Mutat. 1:366-374(1992).

[14]

RP VARIANT MFS SER-2307.

RA MEDLINE=92235290; PubMed=1569206;

RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr.,

RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;

RT "Marfan phenotype variability in a family segregating a missense

RT mutation in the epidermal growth factor-like motif of the fibrillin

RT gene.";

RL J. Clin. Invest. 89:1674-1680(1992).

[15]

RP VARIANTS MFS ILE-548 AND ALA-723.

RA MEDLINE=94010946; PubMed=8406497;

RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,

RA Pyeritz R.E., Francomano C.A.;

RT "Four novel FBNI mutations: significance for mutant transcript level

RT and EGF-like domain calcium binding in the pathogenesis of Marfan

RT syndrome.";

RL Genomics 17:468-475(1993).

[16]

RP VARIANTS MFS SER-2144.

RA MEDLINE=93278402; PubMed=8504310;

RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;

RT "A novel fibrillin mutation in the Marfan syndrome which could

RT disrupt calcium binding of the epidermal growth factor-like module.";

RL Hum. Mol. Genet. 2:475-477(1993).

[17]

RP VARIANTS MFS R-862; Y-1117; P-1137 AND F-1589, AND VARIANT A-1148.

RA MEDLINE=94108431; PubMed=8281141;



RA Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,  
 RA Berg M.A., Miller D.C., Francke U.;  
 RT "Mutation screening of complete fibrillin-1 coding sequence: report  
 RT of five new mutations, including two in 8-cysteine domains.";  
 RL Hum. Mol. Genet. 2:1813-1821(1993).  
 RN [18]  
 RP VARIANTS MFS GLY-217 AND ARG-2627.  
 RX MEDLINE=95067970; PubMed=7977366;  
 RA Karttunen L., Raghunath M., Loengqvist L., Peltonen L.;  
 RT "A compound-heterozygous Marfan patient: two defective fibrillin  
 RT alleles result in a lethal phenotype.";  
 RL Am. J. Hum. Genet. 55:1083-1091(1994).  
 RN [19]  
 RP VARIANT EL LYS-2447.  
 RX MEDLINE=94245249; PubMed=8186302;  
 RA Loengqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,  
 RA Peltonen L.;  
 RT "A novel mutation of the fibrillin gene causing ectopia lentis.";  
 RL Genomics 19:573-576(1994).  
 RN [20]  
 RP VARIANT MFS CYS-627.  
 RX MEDLINE=94272487; PubMed=8004112;  
 RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;  
 RT "Two novel mutations and a neutral polymorphism in EGF-like domains  
 RT of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan  
 RT syndrome patients.";  
 RL Hum. Mol. Genet. 3:373-375(1994).  
 RN [21]  
 RP VARIANT MFS CYS-122.  
 RX MEDLINE=94314977; PubMed=8040326;  
 RA Stahl-Hallengren C., Ukkonen T., Kainulainen K., Kristofersson U.,  
 RA Savne T., Tornqvist K., Peltonen L.;  
 RT "An extra cysteine in one of the non-calcium-binding epidermal growth  
 RT factor-like motifs of the FBN1 polypeptide is connected to a novel  
 RT variant of Marfan syndrome.";  
 RL J. Clin. Invest. 94:709-713(1994).  
 RN [22]  
 RP VARIANT MFS TYR-1223.  
 RX MEDLINE=94351682; PubMed=8071963;  
 RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;  
 RT "A new missense mutation of fibrillin in a patient with Marfan  
 RT syndrome.";  
 RL J. Med. Genet. 31:338-339(1994).  
 RN [23]  
 RP VARIANT MFS HIS-1170.  
 RX MEDLINE=95174777; PubMed=7870075;  
 RA Hayward C., Porteous M.E.M., Brock D.J.H.;  
 RT "A novel mutation in the fibrillin gene (FBN1) in familial  
 RT arachnodactyly.";  
 RL Mol. Cell. Probes 8:325-327(1994).  
 RN [24]  
 RP VARIANTS MFS G-217; N-1023; R-1074; Y-1242; R-1513; E-2127; W-2151;  
 RP K-2447 AND R-2511.  
 RX MEDLINE=94184368; PubMed=8136837;  
 RA Kainulainen K., Karttunen L., Puhakka L., Sakai L., Peltonen L.;  
 RT "Mutations in the fibrillin gene responsible for dominant ectopia  
 RT lentis and neonatal Marfan syndrome.";  
 RL Nat. Genet. 6:64-69(1994).  
 RN [25]  
 RP VARIANT SER-1127.  
 Query Match 13.4%; Score 268.5; DB 1; Length 2871;  
 Best Local Similarity 25.6%; Pred. No. 2,3e-12;  
 Matches 104; Conservative 31; Mismatches 122; Indels 149; Gaps 24;

OY 130 -PGTYGPDCLACOG-----GSORPCSGNGHC-----SGD-GSRQGDGSCRHMGYOG 175  
 DB 1257 IPGER--KCLCYDGRMASBDMKTCYDVMECDLNPVCLSGTEFNKRGSPICGDMGYSK 1314  
 OY 176 L---CTDCMDGYESSLNRNTHSICCTACDESCCTCSGLTRNDCGCEYGVWVLDGACVDV 231  
 DB 1315 KGKTCCTDINECEICAHNCGKHAIVCTNAGSPK-CS-----CSPGWIDGJKICDYL 1364  
 OY 232 DECAAEPPPCSAORCKNANGSYTC-----ECC-----DSCVCGT 267  
 DB 1365 DECSNGTHMCSOHADCKNTMSYRCLCKEGYTGDFCTDLDCESENILCGNGCCLNAP 1424  
 OY 268 G-----EGP-----GNCK-----ECISGYA--REHQ 287  
 DB 1425 GGYRCECMGFVPSADGACEDIDECSLPNICVFETCHNLPGLFCGEIEIGELDRSGN 1484  
 OY 288 CADVDECSLAECTVARKNENCYNPGSYVVCYCPDGF--ETEDACV 331  
 DB 1485 CTDVNEC-LDPTTCI--SGNCVNTPGSYICDCPPDFELNPFVGCV 1527  
 RESULT 5  
 PKCS\_MOUSE STANDARD: PRT; 1877 AA.  
 AC 004592; 062040;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE (Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
 DE (Protein convertase PCS) (Subtilisin/kexin-like protease PCS)  
 DE (Convertase PC5) (PC6) (Subtilisin-like protein convertase 6)  
 DE (SPC6).  
 GN PCSK5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RP [1]  
 RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PCSB).  
 RC STRAIN=ICR; TISSUE=Intestine;  
 RX MEDLINE=93327934; PubMed=8335106;  
 RA Nakagawa T., Murekami K., Nakayama K.;  
 RT "Identification of an isoform with an extremely large Cys-rich region  
 RT of PC6, a Kex2-like processing endoprotease.";  
 RL FEBS Lett. 327:165-171(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM PCSA).  
 RC TISSUE=Brain, and Intestine;  
 RX MEDLINE=93224489; PubMed=8468318;  
 RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murekami K.,  
 RA Nakayama K.;  
 RT "Identification and functional expression of a new member of the  
 RT mammalian Kex2-like processing endoprotease family: its striking  
 RT structural similarity to PC6R4.";  
 RL J. Biochem. 113:132-135(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM PCSA).  
 RC TISSUE=Adrenal cortex;  
 RX MEDLINE=93342056; PubMed=8341687;  
 RA Lussan J., Vileau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
 RT "CDNA structure of the mouse and rat subtilisin/kexin-like PC5: a  
 RT candidate protein convertase expressed in endocrine and  
 RT nonendocrine cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
 RN [4]  
 RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=97103178; PubMed=8947550;  
 RA De Bie I., Marchlewicz M., Malide D., Lazure C., Nakayama K.,  
 RA Bendayan M., Seidah N.G.;  
 RT "The isoforms of proprotein convertase PC5 are sorted to different  
 RT subcellular compartments.";  
 RL J. Cell Biol. 135:1261-1275(1996).  
 RN [5]







```

FT CARBOHYD 537 537 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 541 541 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 569 705 PROKEDPTVRCISCRNDACVADPHYTHYSIPTR
FT EFTREKEDPTVRCISCRNDACVADPHYTHYSIPTR
FT KRYEDGNTVGVQVQVPRVGFVAVLKLENNYVGVSHR
FT NVVNVHVFSEYWF -> RCARLPCHEOCPRPLRTIY
FT HLESPNIOGVPAVFERMGSSAVDGDMOLATIRNEGEFF
FT TTRKVSHGSCVALTRPIPEPRDILLTKVMDIVRHGTVSSEF
FT VAKLFTVSNEL (IN ISOFOR D).
SQ SEQUENCE 705 AA: 78056 MW: FD3F06469A4BAE2B CRC64:

Query Match 12.6%: Score 253.5; DB 1; Length 705;
Best Local Similarity 27.1%: Pred. No. 8.3e-12;
Matches 80; Conservative 24; Mismatches 108; Indels 83; Gaps 14;

OY 115 LFEFVCVTKKCCSPGTYGPDCLACOGS-QRP-----CSGNHGCSG 156
DB 131 MISYQGLVFRACCVKARENSDFVQGNQADLPDAKIPDEEDQDPYLNDRCGGSPCKQ 190
OY 157 DGSNQGSG-SCRGHMGVQ-----GRLCTD---CMGYFS-----SLRNETHICTACDESC 203
DB 191 QCRDPTGVETICSCFVGYQLQSDGSCEDINECTGSHNCRLGESCINTVGSFRCQROSSC 250
OY 204 KTCGSLTN---RDGCECEV-----GWVLDE-GACVD 230
DB 251 GTGEVLEEDNNCKRIDECETGHCNCPDPTFCNTLGRPCRPKLOCKSGFIQDALGNCTID 310
OY 231 VDECAEPPEPSAAQFCRNANGSYTCECDSSCVGCTGEGPNCKECISG-AREHG-QC 288
DB 311 INECLTISAPCPVQOTCINTEGSYTCOK-----NVPCNGRGYHLNEEGTRC 356
OY 289 ADVDECSLAETCYCRKNKNCNTPGSYVCVCPDG--FEETEDACVPPAEATGEG 341
DB 357 VDVECAPPAEPC-GKGHHCLNSPSCFCECKAGFYFDGISRTGVINECQRYPG 410

RESULT 8
FBN1_MOUSE STANDARD; PRT: 2871 AA.
ID FBN1_MOUSE STANDARD; PRT: 2871 AA.
AC Q61534; Q60826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130561; Pubmed=7829516;
RA Yin W., Germiller J., Sanguinetti C., Smiley E., Panglillian T.,
RA Pareira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene.";
RL J. Biol. Chem. 270:1798-1806(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Ota K., Kumar A., Mada J., Liu Z., Kanwar Y.S.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: L29454; AA556840.1; -
DR EMBL: U22493; AA64217.1; -
DR HSSP: P35555; IAPJ.
DR MGD: MGI:95489; Fbn1.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002212; TB.
DR Pfam: PF00008; EGF; 46.
DR Pfam: PF00683; TB; 9.
DR SMART: SM00179; EGF_Ca; 42.
DR SMART: SM00001; EGF_Like; 4.
DR PROSITE: PS00010; ASX_HYDROXYL; 43.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 38.
DR PROSITE: PS01187; EGF_Ca; 45.
DR Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27
FT CHAIN 28 2871
FT DOMAIN 81 112
FT DOMAIN 115 146
FT DOMAIN 147 178
FT DOMAIN 246 287
FT DOMAIN 288 329
FT DOMAIN 330 401
FT DOMAIN 402 446
FT DOMAIN 449 489
FT DOMAIN 490 529
FT DOMAIN 530 571
FT DOMAIN 572 612
FT DOMAIN 613 653
FT DOMAIN 723 764
FT DOMAIN 765 806
FT DOMAIN 807 846
FT DOMAIN 910 951
FT DOMAIN 952 1018
FT DOMAIN 1028 1069
FT DOMAIN 1070 1112
FT DOMAIN 1113 1154
FT DOMAIN 1155 1196
FT DOMAIN 1197 1237
FT DOMAIN 1238 1279
FT DOMAIN 1280 1321
FT DOMAIN 1322 1362
FT DOMAIN 1363 1403
FT DOMAIN 1404 1445
FT DOMAIN 1446 1486
FT DOMAIN 1487 1527
FT DOMAIN 1528 1568
FT DOMAIN 1569 1608
FT DOMAIN 1609 1647
FT DOMAIN 1648 1688
FT DOMAIN 1689 1728
FT DOMAIN 1766 1807
FT DOMAIN 1808 1848
FT DOMAIN 1849 1890
FT DOMAIN 1891 1929
FT DOMAIN 1930 1972
FT DOMAIN 1973 2012
FT DOMAIN 2013 2054
FT DOMAIN 2055 2121
FT DOMAIN 2127 2165
FT DOMAIN 2166 2205
FT DOMAIN 2206 2246

```

```

FT DOMAIN 2247 2290 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2291 2332 EGF-LIKE 40, CALCIUM-BINDING.
FT DOMAIN 2333 2400 TGFBR 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 89 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 160 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
FT DISULFID 273 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 790 796 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.

```

```

FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.

```

## Query Match

12.6% Score 252.5; DB 1; Length 2871;

Best Local Similarity 25.1%; Pred No. 3.6e-11; Matches 102; Conservative 31; Mismatches 124; Indels 149; Gaps 24;

```

QY 30 PCHRCRGLVDFK-----NOGMVDTAKRFGGNTAMEKTLISKYSESEIRLLEILGCLCES 85
DB 1167 PHRCVNLGKYGQACACNPVHPTDRLF-----CYDIDECSIMNGCET 1210
QY 86 -----SDPECN-----QMLEAOEHLKAWMLQKSEYDLEFEMFCVTKLKYC-----CS- 129
DB 1211 FCTNSDGSYECSCQPGFALMPDR-----SCRDIDQ--CEDNPNICDGGQCTN 1256
QY 130 -PETYGPDLAOG-----GSORPCSGNGHC-----SGD--GSROGDSCRCRMGYOGP 175
DB 1257 IPEGF--RCLCYDGFMASEDMKTCVYNNECDLNPATCLSGTCENKRGSTICHCMDMYSK 1314
QY 176 L-----CTDCMDGYFSSLRNETHICTACDESCCTCGSLNRRDGCCEGVWLDGACVDV 231
DB 1315 KGKTGCTDINECHIGAHNGCRHAYCTNTAGSFK-CS-----CSPGMIGGKICTDL 1364
QY 232 DECAAPPPCSAOPCKNANGSYTC-----EBC-----DSSVYCGT 267
DB 1365 DECSNGTHMCSOHADCKNTMGSRCLCKDGYTGDTCTDDECSBNLNLGCGOQLNAP 1424
QY 268 G-----GEP-----GNCK-----ECISGYA--REHQ 287
DB 1425 GGYRCECDMGFVPSADCKACEDIDECSLNICYFGCHNLPGLFRCECEIGYELDRSGCN 1484
QY 288 CADVDECSLAETCVKRNENYTPGSIYVCVCPDGE--ETEDACV 331
DB 1485 CTDVNEC-LDPTTCI--SGNCVTPGSIYCDSPDELNPTRVGCY 1527

RESULT 9
ID FBN2_MOUSE STANDARD; PRT; 2907 AA.
AC 061555; 063957;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Fibrillin 2 precursor.
GN FBN2 OR FBN-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95263670; PubMed=7744963;
RT Zhang H., Hu W., Ramirez F.;
of extracellular microfibrils."

```





```

FT DISULFID 1112 1124 BY SIMILARITY.
FT DISULFID 1119 1133 BY SIMILARITY.
FT DISULFID 1135 1149 BY SIMILARITY.
FT DISULFID 1155 1167 BY SIMILARITY.
FT DISULFID 1162 1176 BY SIMILARITY.
FT DISULFID 1178 1191 BY SIMILARITY.
FT DISULFID 1197 1209 BY SIMILARITY.
FT DISULFID 1204 1218 BY SIMILARITY.
FT DISULFID 1220 1233 BY SIMILARITY.
FT DISULFID 1239 1250 BY SIMILARITY.
FT DISULFID 1246 1259 BY SIMILARITY.
FT DISULFID 1261 1274 BY SIMILARITY.
FT DISULFID 1280 1292 BY SIMILARITY.
FT DISULFID 1287 1301 BY SIMILARITY.
FT DISULFID 1303 1316 BY SIMILARITY.
FT DISULFID 1322 1334 BY SIMILARITY.
FT DISULFID 1329 1343 BY SIMILARITY.
FT DISULFID 1345 1358 BY SIMILARITY.
FT DISULFID 1364 1377 BY SIMILARITY.
FT DISULFID 1371 1386 BY SIMILARITY.
FT DISULFID 1388 1399 BY SIMILARITY.
FT DISULFID 1405 1418 BY SIMILARITY.
FT DISULFID 1412 1427 BY SIMILARITY.
FT DISULFID 1429 1440 BY SIMILARITY.
FT DISULFID 1446 1458 BY SIMILARITY.
FT DISULFID 1453 1467 BY SIMILARITY.
FT DISULFID 1469 1482 BY SIMILARITY.
FT DISULFID 1488 1499 BY SIMILARITY.
FT DISULFID 1494 1508 BY SIMILARITY.
FT DISULFID 1510 1523 BY SIMILARITY.
FT DISULFID 1529 1540 BY SIMILARITY.
FT DISULFID 1535 1549 BY SIMILARITY.
FT DISULFID 1551 1564 BY SIMILARITY.
FT DISULFID 1647 1659 BY SIMILARITY.
FT DISULFID 1654 1668 BY SIMILARITY.
FT DISULFID 1670 1683 BY SIMILARITY.
FT DISULFID 1689 1701 BY SIMILARITY.

```

Query Match 12.6%; Score 252.5; DB 1; Length 2907;  
 Best Local Similarity 24.8%; Pred. No. 3.7e-11;  
 Matches 103; Conservative 30; Mismatches 123; Indels 159; Gaps 25;

```

OY 23 EAAKPTPC-HHCRGLVDFKNOGAVDTAKNFGGNTANEKTLKESSEIRLEIEEG 81
DB 2249 ECAQNPPLCAFRGCMNTFGSYE-CTCPVGYG-LREDQMKCK-DLDECAEG 2294
OY 82 L-CESSDPCNQMLEAOEHL EAMWLDKSEYRDLFEFVCYKTLKVCSPG-TYGPDL 138
DB 2295 LHDCEGRGCMCKNL-IGTFMCICPPGARRPDE 2327
OY 139 ACQGGSQ--RP-CGNGHCSDGSRGDSRCRMGYQGPLC-TDCMDG- 183
DB 2328 GGVDEHCETKPGICE-NRCV--NIISYRCCEGEGSSSGTECLDNOGLCFAY 2383
OY 184 -----YFSLAN-ETHSICTACD- 200
DB 2384 LQTMQMASSSRNLVTKSPC-CCDGRGWHOCLELCPRLGTAYKKICPHGPGYATDGR 2442
OY 201 -ESCTGSGT-TNRDGE-----CEVWTLDEG--ACVVDDECAAPPPCSAAQAF 248
DB 2443 IDECKVMSLCTNGOCVNTMGSPRCFCCKAGYIMDISGTACVDDECSQSPKPCNF--ICK 2500
OY 249 NANGSYTC-----EECDSS-----CVGCTGRGPNCKRECIISGYAR 283
DB 2501 NFKGSYOCSCPRGYVLQEDGKTKDLDECQTKOHNOGLCFVMTLG--GFTCKCPRGFTQ 2557
OY 284 EHGQCADVDEGLAEKTCYRKNNENYNTPGSYVCCPDGFE-----ETDACC 330
DB 2558 HHTACIDNNECGSPSLCGAKG-ICONTGPSFCCEQGRGFSIDASGLNCEVDDEC 2611

```

RESULT 10  
 FBN2\_HUMAN

```

ID FBN2_HUMAN STANDARD: PRT; 2911 AA.
AC P35556;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibrillin 2 precursor.
GN FBN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP MEDLINE=94165150; PubMed=8120105;
RA Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sanguineti C.,
RA Bonadio J., Mecham R.P., Ramirez F.;
RT "Structure and expression of fibrillin-2, a novel microfibrillar
RT component preferentially located in elastic matrices.";
RL J. Cell Biol. 124:855-863(1994).
RN [2]
RP MEDLINE=91304567; PubMed=1852206;
RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA Tsiipouras P., Ramirez F., Hollister D.;
RT "Linkage of Marfan syndrome and a phenotypically related disorder to
RT two different fibrillin genes.";
RL Nature 352:330-334(1991).
RN [3]
RP VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.
RX MEDLINE=96083599; PubMed=7493032;
RA Putnam E.A., Zhang H., Ramirez F., Milewicz D.M.;
RT "Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,
RT congenital contractual arachnoidactyly.";
RL Nat. Genet. 11:456-458(1995).
RN [4]
RP VARIANTS CCA HIS-1114.
RX MEDLINE=98407789; PubMed=9737771;
RA Babcock D., Gasner C., Franke U., Maslan C.;
RT "A single mutation that results in an asp-to-his substitution and
RT partial exon skipping in a family with congenital contractual
RT arachnoidactyly.";
RL Hum. Genet. 103:22-28(1998).
RN [5]
RP VARIANTS CCA PHE-1141 AND TRP-1252.
RX MEDLINE=20259236; PubMed=10797416;
RA Belleh S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,
RA Godfrey M.;
RT "Two novel fibrillin-2 mutations in congenital contractual
RT arachnoidactyly.";
RL Am. J. Med. Genet. 92:7-12(2000).
CC -I- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-2-CONTAINING MICROFIBRILS REGULATE
CC THE EARLY PROCESS OF ELASTIC FIBER ASSEMBLY.
CC -I- DISEASE: DEFECTS IN FBN2 ARE THE CAUSE OF CONGENITAL CONTRACTUAL
CC ARACHNOIDACTYLY (CCA) (ALSO KNOWN AS BEALS SYNDROME). CCA IS
CC PHENOTYPICALLY SIMILAR TO MARFAN SYNDROME, BUT DOES NOT EFFECT THE
CC AORTA AND THE EYES.
CC -I- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U03272; AAA18950.1;
CC EMBL: X62009; -; NOT_ANNOTATED_CDS.
CC PIR: S17063; S17063.
CC PIR: S31101; S31101.

```



DR	HSSP; P35555; IEMN.
DR	MIM; 121050; -;
DR	InterPro; IPRO00152; Asx_hydroxyl.
DR	InterPro; IPRO00561; EGF-like.
DR	InterPro; IPRO01881; EGF_Ca.
DR	InterPro; IPRO01438; EGF_LI.
DR	InterPro; IPRO02212; TB.
DR	Pfam; PF00008; EGF_45.
DR	Pfam; PF00683; TB; 9.
DR	PRINTS; PRO0010; EGFBLOOD.
DR	SMART; SM00179; EGF_Ca; 43.
DR	SMART; SM00001; EGF_like; 3.
DR	PROSITE; PS00010; ASX_HYDROXYL; 43.
DR	PROSITE; PS00022; EGF_L1; 2.
DR	PROSITE; PS0186; EGF_2; 37.
DR	PROSITE; PS0187; EGF_CA; 43.
KW	Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain; Repeat; signal; Multigene family; Disease mutation; Polymorphism.
KW	Repeat; signal; Multigene family; Disease mutation; Polymorphism.
FT	SIGNAL
FT	CHAIN
FT	29 2911
FT	FIBRILIN 2.
FT	DOMAIN
FT	111 142
FT	EGF-LIKE 1, NON-CALCIUM BINDING.
FT	DOMAIN
FT	145 176
FT	EGF-LIKE 2, NON-CALCIUM BINDING.
FT	DOMAIN
FT	176 207
FT	EGF-LIKE 3, NON-CALCIUM BINDING.
FT	DOMAIN
FT	275 316
FT	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN
FT	317 358
FT	EGF-LIKE 5, CALCIUM-BINDING.
FT	REPEAT
FT	359 425
FT	TGEBP 1.
FT	DOMAIN
FT	493 533
FT	EGF-LIKE 6, NON-CALCIUM BINDING.
FT	DOMAIN
FT	534 573
FT	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN
FT	574 615
FT	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN
FT	616 656
FT	EGF-LIKE 9, CALCIUM-BINDING.
FT	DOMAIN
FT	657 697
FT	EGF-LIKE 10, CALCIUM-BINDING.
FT	REPEAT
FT	698 766
FT	TGEBP 2.
FT	DOMAIN
FT	767 808
FT	EGF-LIKE 11, CALCIUM-BINDING.
FT	DOMAIN
FT	809 850
FT	EGF-LIKE 12, CALCIUM-BINDING.
FT	DOMAIN
FT	851 890
FT	EGF-LIKE 13, CALCIUM-BINDING.
FT	DOMAIN
FT	954 995
FT	EGF-LIKE 14, CALCIUM-BINDING.
FT	REPEAT
FT	996 1071
FT	TGEBP 3.
FT	DOMAIN
FT	1072 1113
FT	EGF-LIKE 15, CALCIUM-BINDING.
FT	DOMAIN
FT	1114 1156
FT	EGF-LIKE 16, CALCIUM-BINDING.
FT	DOMAIN
FT	1157 1198
FT	EGF-LIKE 17, CALCIUM-BINDING.
FT	DOMAIN
FT	1199 1240
FT	EGF-LIKE 18, CALCIUM-BINDING.
FT	DOMAIN
FT	1241 1281
FT	EGF-LIKE 19, CALCIUM-BINDING.
FT	DOMAIN
FT	1282 1323
FT	EGF-LIKE 20, CALCIUM-BINDING.
FT	DOMAIN
FT	1324 1365
FT	EGF-LIKE 21, CALCIUM-BINDING.
FT	DOMAIN
FT	1366 1406
FT	EGF-LIKE 22, CALCIUM-BINDING.
FT	DOMAIN
FT	1407 1447
FT	EGF-LIKE 23, CALCIUM-BINDING.
FT	DOMAIN
FT	1448 1489
FT	EGF-LIKE 24, CALCIUM-BINDING.
FT	DOMAIN
FT	1490 1530
FT	EGF-LIKE 25, CALCIUM-BINDING.
FT	DOMAIN
FT	1531 1571
FT	EGF-LIKE 26, CALCIUM-BINDING.
FT	REPEAT
FT	1572 1648
FT	TGEBP 4.
FT	DOMAIN
FT	1649 1690
FT	EGF-LIKE 27, CALCIUM-BINDING.
FT	DOMAIN
FT	1691 1732
FT	EGF-LIKE 28, CALCIUM-BINDING.
FT	REPEAT
FT	1733 1806
FT	TGEBP 5.
FT	DOMAIN
FT	1807 1848
FT	EGF-LIKE 29, CALCIUM-BINDING.
FT	DOMAIN
FT	1849 1890
FT	EGF-LIKE 30, CALCIUM-BINDING.
FT	DOMAIN
FT	1891 1932
FT	EGF-LIKE 31, CALCIUM-BINDING.
FT	DOMAIN
FT	1933 1971
FT	EGF-LIKE 32, CALCIUM-BINDING.
FT	DOMAIN
FT	1972 2014
FT	EGF-LIKE 33, CALCIUM-BINDING.
FT	DOMAIN
FT	2015 2054
FT	EGF-LIKE 34, CALCIUM-BINDING.
FT	DOMAIN
FT	2055 2096
FT	EGF-LIKE 35, CALCIUM-BINDING.
FT	REPEAT
FT	2097 2169
FT	TGEBP 6.
FT	DOMAIN
FT	2170 2211
FT	EGF-LIKE 36, CALCIUM-BINDING.
FT	DOMAIN
FT	2212 2251
FT	EGF-LIKE 37, CALCIUM-BINDING.
FT	DOMAIN
FT	2252 2292
FT	EGF-LIKE 38, CALCIUM-BINDING.
FT	DOMAIN
FT	2293 2336
FT	EGF-LIKE 39, CALCIUM-BINDING.
FT	DOMAIN
FT	2337 2378
FT	EGF-LIKE 40, CALCIUM-BINDING.
FT	REPEAT
FT	2379 2447
FT	TGEBP 7.
FT	DOMAIN
FT	2448 2489
FT	EGF-LIKE 41, CALCIUM-BINDING.
FT	DOMAIN
FT	2490 2530
FT	EGF-LIKE 42, CALCIUM-BINDING.
FT	DOMAIN
FT	2531 2569
FT	EGF-LIKE 43, CALCIUM-BINDING.
FT	DOMAIN
FT	2570 2612
FT	EGF-LIKE 44, CALCIUM-BINDING.
FT	DOMAIN
FT	2613 2652

QY	152	GHCSGDSGSGRCHGMY	12.4% Best Local Similarity 21.1%; Matches 100; Conservative 53; Mismatches 125; Indels 197; Gaps 26	Score 249.5; DB 1; Length 2911; Pred. No. 6; 2e-11;	EGF-Like 47, CALCIUM-BINDING.
FT	DOMAIN	2654	2733	EGF-Like 47, CALCIUM-BINDING.	
FT	DISULFID	115	124	BY SIMILARITY.	
FT	DISULFID	119	130	BY SIMILARITY.	
FT	DISULFID	132	141	BY SIMILARITY.	
FT	DISULFID	149	159	BY SIMILARITY.	
FT	DISULFID	153	164	BY SIMILARITY.	
FT	DISULFID	166	175	BY SIMILARITY.	
FT	DISULFID	180	190	BY SIMILARITY.	
FT	DISULFID	184	195	BY SIMILARITY.	
FT	DISULFID	197	206	BY SIMILARITY.	
FT	DISULFID	279	291	BY SIMILARITY.	
FT	DISULFID	286	300	BY SIMILARITY.	
FT	DISULFID	302	315	BY SIMILARITY.	
FT	DISULFID	321	333	BY SIMILARITY.	
FT	DISULFID	328	342	BY SIMILARITY.	
FT	DISULFID	344	357	BY SIMILARITY.	
FT	DISULFID	497	509	BY SIMILARITY.	
FT	DISULFID	504	518	BY SIMILARITY.	
FT	DISULFID	520	532	BY SIMILARITY.	
FT	DISULFID	538	548	BY SIMILARITY.	
FT	DISULFID	543	557	BY SIMILARITY.	
FT	DISULFID	559	572	BY SIMILARITY.	
FT	DISULFID	578	590	BY SIMILARITY.	
FT	DISULFID	585	599	BY SIMILARITY.	
FT	DISULFID	601	614	BY SIMILARITY.	
FT	DISULFID	620	631	BY SIMILARITY.	
FT	DISULFID	626	640	BY SIMILARITY.	
FT	DISULFID	642	655	BY SIMILARITY.	
FT	DISULFID	661	672	BY SIMILARITY.	
FT	DISULFID	667	681	BY SIMILARITY.	
FT	DISULFID	683	696	BY SIMILARITY.	
FT	DISULFID	771	783	BY SIMILARITY.	
FT	DISULFID	778	792	BY SIMILARITY.	
FT	DISULFID	794	807	BY SIMILARITY.	
FT	DISULFID	813	825	BY SIMILARITY.	
FT	DISULFID	820	834	BY SIMILARITY.	
FT	DISULFID	836	849	BY SIMILARITY.	
FT	DISULFID	855	865	BY SIMILARITY.	
FT	DISULFID	860	874	BY SIMILARITY.	
FT	DISULFID	876	889	BY SIMILARITY.	
FT	DISULFID	958	970	BY SIMILARITY.	
FT	DISULFID	965	979	BY SIMILARITY.	
FT	DISULFID	981	994	BY SIMILARITY.	
FT	DISULFID	1076	1088	BY SIMILARITY.	
FT	DISULFID	1083	1097	BY SIMILARITY.	
FT	DISULFID	1099	1112	BY SIMILARITY.	
FT	DISULFID	1118	1130	BY SIMILARITY.	
FT	DISULFID	1125	1139	BY SIMILARITY.	
FT	DISULFID	1141	1155	BY SIMILARITY.	
FT	DISULFID	1161	1173	BY SIMILARITY.	
FT	DISULFID	1168	1182	BY SIMILARITY.	
FT	DISULFID	1184	1197	BY SIMILARITY.	
FT	DISULFID	1203	1215	BY SIMILARITY.	
FT	DISULFID	1210	1224	BY SIMILARITY.	

```

Db 1260 -----EGSYECSSSEGYALAMPDGRSCADIDECENNPDI CDGAGCTINIGEXRCLCY 1310
Qy 182 DGFESLNRHETHISICTACDESKTC-----SGLTN-RDCGEC 217
Db 1311 DGFMA5MDMKTCITDYNEDCLNSITCMFEGECENTKGSFICHCQLGYSVKRGTTGCDYVDEC 1370
Qy 218 EV-----GWLVDGACVVDVDECAAEPPPCSAQFCNKANS 253
Db 1371 ELGAINCMHMASCLINPQSFKSCBEGMIGNIKICIDIDECNNGHQSINACQVNTPGS 1430
Qy 254 YTCEDSSSCVCGCTEG-----PGNCK-ECISGY--AREHG 286
Db 1431 YRC-ACSE---GFTGDGFTCSVDVDECAENINLCENGQCLNVPAGVRCBCEMEGFTPADSR 1486
Qy 287 OCADVDECSLAERTYVRKNENCYNTPGSYVCYCPGFEETE-----DACVP 333
Db 1487 SCQDIDECSF-ONICV--SGTCNNNPGMFGHICIDGVEIDRTGNGCTDIDECADP 1538

RESULT 11
PKS5_BRACL
ID PKS5_BRACL STANDARD; PRT: 1696 AA.
AC Q9NJ15; Q9NJ16; Q9NJ14;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Proprotein convertase PC6-1like) (apc6).
GN PC6.
OS Brachyostoma californiensis (Calliophoria lancelet) (Amphioxus).
OC Eumetazoa; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7738;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RX MEDLINE=20175281; PubMed=10708868;
RA Oliva A.A. Jr., Chan S.J., Steiner D.F.;
RT "Evolution of the prohormone convertases: identification of a
RL homologue of PC6 in the protochordate amphioxus."
RL Biochim. Biophys. Acta 1477:338-348(2000).
CC 1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC 1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
CC 1- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
CC 1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC 1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.
CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC 1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF184615; AAF26300.1; -
DR EMBL; AF184616; AAF26301.1; -
DR EMBL; AF184617; AAF26302.1; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002884; P_domain.
DR InterPro; IPR000209; Peptidase_S8.
DR PRINTS; PR00723; SUBTILISIN.

```

DR	PriDom:	PD000717; P.domain; 1.
DR	SMART;	SM00184; EGF; 2.
DR	SMART;	SM00261; FU; 17.
DR	PROSITE;	PS00136; SUBTILASE_ASP; FALSE_NEG.
DR	PROSITE;	PS00137; SUBTILASE_HIS; 1.
DR	PROSITE;	PS00138; SUBTILASE_SERP; 1.
KW	Hydrolase:	Serine protease; Glycoprotein; Zymogen; Signal;
KW	Cleavage on pair of basic residues;	Repeat; Alternative splicing;
KW	Transmembrane.	
FT	SIGNAL	1 25 POTENTIAL.
FT	PROPEP	26 110 POTENTIAL.
FT	CHAIN	111 1696 PROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 5.
FT	DOMAIN	111 1618 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1619 1639 POTENTIAL.
FT	DOMAIN	1640 1696 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	111 488 CATALYTIC.
FT	DOMAIN	496 637 HOMO B.
FT	DOMAIN	664 1649 CYS-RICH MOTIF (CRM) REGION.
FT	SITE	110 111 CLEAVAGE (AUTO-) (BY SIMILARITY).
FT	ACT_SITE	192 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD	246 246 N-LINKED (GLCNAC . . . ) (POTENTIAL).
FT	CARBOHYD	529 529 N-LINKED (GLCNAC . . . ) (POTENTIAL).
FT	CARBOHYD	885 885 N-LINKED (GLCNAC . . . ) (POTENTIAL).
FT	VARSPLIC	1259 1323 DDTIADRGECITSCPGEMRREKRCACHPTCKESDSEY DDTCACNDGFIITLDASSCEAGCP -> AENONASFCPPA PREVSVALIELALGHRLSYLTIDVPPQSNSSPDTVLGNADRARL TTANSAGRA (IN ISOFORM C).
FT	VARSPLIC	1324 1696 MISSING (IN ISOFORM C).
FT	VARSPLIC	1288 1343 CHPTCKESDSEYDDTCACNDGFIITLDASSCEAGCPGFL HHGGDSCCHRECKT -> IARCVRDRORSWCMDLVLRNFVC VRRYPVKRCOCCTCKLYMEDRPMRRSSOPTOGS (IN ISOFORM A).
FT	VARSPLIC	1344 1696 MISSING (IN ISOFORM A).
FT	SEQUENCE	1696 AA; 188410 MW; 281CBEL1784257CBD CRC64;
SO	SEQUENCE	
Query Match		12.4%; Score 248.5; DB 1; Length 1696;
Best Local Similarity		19.9%; Pred. No. 4,4e-11;
Matches		78; Conservative 40; Mismatches 101; Indels 173; Gaps 16
QY	89 ECNMQLF-AOEHLIAWMWLKSEYPDLEFMFYCYTKLAVCCSPGTYG-----P	135
DB	938 QCNQCITPYEGEDHF-----LYGSTC---HVTCPGLIGDITDOVCKACAP	980
QY	136 DCLAQGGSGQRCS-----GNHGCSDSGRQDGSR--CHMKYORP-----	175
DB	981 GTCIADGPADNOCITCEEERAPTDNRCSQSQTDEAFCAECSCHSCEEPDI CDSD EBY	1040
QY	176 -----LCITDCMDFESS-----	187
DB	1041 YITEDTCYARTNCPSFTYRDDDDRECRPCPHDNCEACDNGNNANCNSCKRGFKTPPGCGST	1100
QY	188 -----LNETHHSICTACDESGTSGLLTNRDGCEGEVWVLDEGAC-----VD	230
DB	1101 GCPRNYKYDDTKKECKPCDSCFCTSGPASFFCLSCADDGDLFHSSCRSTCPAGFIGNAE	1160
QY	231 VDECAAEPPPGCAAFPCKANANGSYICEBCDSSCVGTGTGGPGNKCEICIGYAREHQCAD	290
DB	1161 SHECVES--SCBODDYSETG--RCEDCPYNCRCAC--DNDDCAECAAPTYYIVDGRCRP	1214
QY	291 VDEC-----SLAETCYRKHNKCNT-----P	312
DB	1215 ETCEEDGEYQDRDRDTALIESCRPHQSCSTGSGPSDITDCDSCKGDPTIIDREGCITSCGP	1274
QY	313 GSIV-----C-VCPDGFEEETDAC	330
DB	1275 GEYMDRREKKCKACHPTCKESCSEYDDITCTAC	1306

PRS-RAT  
 ID PRS-RAT STANDARD: PRT: 675 AA.  
 AC P53813;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Vitamin K-dependent protein S precursor.  
 GN PROS1 OR PROS.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID-10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Liver;  
 RA MEDLINE-9533263; PubMed-7608128;  
 RT Yasuda F., Haysht T., Tanitame K., Nishioka J., Suzuki K.;  
 RT "Molecular cloning and functional characterization of rat plasma  
 protein S.";  
 RL J. Biochem. 117:374-383(1995).  
 CC - FUNCTION: ANTICOAGULANT PLASMA PROTEIN; IT IS A COFACTOR TO  
 CC ACTIVATED PROTEIN C IN THE DEGRADATION OF COAGULATION FACTORS VA  
 CC AND VIIIA. IT HELP TO PREVENT COAGULATION AND STIMULATING  
 CC FIBRINOLYSIS.  
 CC - SUBCELLULAR LOCATION: Extracellular.  
 CC - TISSUE SPECIFICITY: PLASMA.  
 CC - SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.  
 CC - SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: S78744; AAC60704.1; -  
 DR HSSP: P00740; 1CFH.  
 DR InterPro: IPR000153; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR002383; GLA\_blood.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR000294; VitK\_dep\_GLA.  
 DR Pfam: PF00008; EGF: 4.  
 DR Pfam: PF00594; gla: 1.  
 DR Pfam: PF00054; laminin\_G: 1.  
 DR PRINTS: PRO0001; GLABLOOD.  
 DR SMART: SM00179; EGF\_CA: 3.  
 DR SMART: SM00001; EGF\_Like: 1.  
 DR SMART: SM00069; GLA: 1.  
 DR SMART: SM00282; LamG: 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL: 4.  
 DR PROSITE: PS00022; EGF\_1: 1.  
 DR PROSITE: PS0186; EGF\_2: 3.  
 DR PROSITE: PS0187; EGF\_CA: 3.  
 DR PROSITE: PS00011; GLU CARBOXYLATION: 1.  
 DR PROSITE: PS00025; LAM\_G\_DOMAIN: 1.  
 KW Plasma; Gamma-carboxyglutamic acid; Calcium; Vitamin K; Zymogen;  
 KW Blood coagulation; Hydroxylation; Glycoprotein; Signal; Repeat;  
 KW EGF-like domain.  
 FT SIGNAL 1 24  
 FT PROPEP 25 41 BY SIMILARITY.  
 FT CHAIN 42 675 BY SIMILARITY.  
 FT DOMAIN 42 675 VITAMIN K-DEPENDENT PROTEIN S.  
 FT DOMAIN 87 86 GLA.  
 FT DOMAIN 87 116 THROMBIN-SENSITIVE.  
 FT DOMAIN 117 155 EGF-LIKE 1.  
 FT DOMAIN 157 200 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 201 242 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 243 283 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 299 475 LAMININ G-LIKE 1.  
 FT DOMAIN 484 665 LAMININ G-LIKE 2.

FT MOD\_RES 47 47 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 48 48 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 55 55 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 57 57 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 60 60 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 61 61 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 66 66 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 67 67 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 70 70 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 73 73 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 77 77 GAMMA-CARBOXYGLUTAMIC ACID  
 FT (BY SIMILARITY).  
 FT MOD\_RES 136 136 HYDROXYLATION (BY SIMILARITY).  
 FT DISULFID 121 134 BY SIMILARITY.  
 FT DISULFID 126 143 BY SIMILARITY.  
 FT DISULFID 145 154 BY SIMILARITY.  
 FT DISULFID 161 175 BY SIMILARITY.  
 FT DISULFID 171 184 BY SIMILARITY.  
 FT DISULFID 186 199 BY SIMILARITY.  
 FT DISULFID 205 217 BY SIMILARITY.  
 FT DISULFID 212 226 BY SIMILARITY.  
 FT DISULFID 228 241 BY SIMILARITY.  
 FT DISULFID 247 256 BY SIMILARITY.  
 FT DISULFID 252 265 BY SIMILARITY.  
 FT DISULFID 267 282 BY SIMILARITY.  
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 509 509 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 675 AA; 74626 MW; B4338F756BA86075 CRC64;

Query Match 12.38; Score 247; DB 1; Length 675;  
 Best Local Similarity 26.98; Pred. No. 2,4e-11;  
 Matches 93; Conservative 34; Mismatches 113; Indels 106; Gaps 21;  
 11 LPLPLLLPAPPAEAKKPTPCRCGLVDKRNQGMVDTAKNFGGNTAMEKTLKTES 70  
 11 LLAALAVLPNSET-----NPLSKERASQVLRKR--ANTLLETRKKNLER 56  
 71 SEIRLEILLEGICSDPECNOMLEAOEH-----LEAWML-----OLKSEYP 113  
 57 -----ECIEELCNKE--EAREVEENNPEDYFPKYLGLGARRVGAFAARSANAYP 108  
 114 DLFEMFCVKTLYKVCSPGTGPDCLACQGGSGRPGSCNGHCS-GDGSROGDGSCRGHMGY 172  
 109 DLRS--CVNALPDDCDP-----MPCNEDGYLSCKDG--QGAFCTICKPGW 149  
 173 QGRPLC-----TDCKMGYSSLRNETHSICACDECKTCSGLTNDGCEVEYGMVL--DEG 226  
 150 QGDGCGQPDINECKD-----PSNINGGCSQTCDDTPG--SYHC--SKIGFAMLRNKK 197  
 227 ACVDVDECAAPPCSAAPFCNKNANGSYTECECSCVCGTGEGPGNCKEICISYAREHG 286  
 198 DCKRYDECSLKPSCVGR-VCKNIPGDEFC-----ECPNRIYRDP 237  
 287 --OCADVDECSLAERTCVRNKNENYNTPGSYVCV--PDGEETED 328  
 238 SKSCGDVDECS--ENTCA--QLCVNYPGGSYCYCDKKKGKLAOD 278  
 RESULT 13  
 FB12\_MOUSE  
 ID FB12\_MOUSE STANDARD: PRT: 1221 AA.  
 AC P37889; Q9W012;

```

DE 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE fibulin-2 precursor.
GN FBLN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090.
RN
RP [1]
RN
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-35.
RC TISSUE=Fibroblast;
RX MEDLINE=94064787; PubMed=8245130;
RA Graessels S., Sicot F.-X., Gotta S., Chu M.-L.;
RT "Mouse fibulin-2 gene. Complete exon-intron organization and promoter
RT characterization."
RL Eur. J. Biochem. 263:471-477(1999).
CC -1- FUNCTION: ITS BINDING TO FIBRONECTIN AND SOME OTHER LIGANDS IS
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS OF FIBULIN-2 DUE TO THE
CC PRESENCE OR ABSENCE OF A SINGLE EGF-LIKE (3) REPEAT MAY ARISE BY
CC ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: COMPONENT OF BOTH BASEMENT MEMBRANES AND OTHER
CC CONNECTIVE TISSUES.
CC -1- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 11 EGF-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
DR EMBL; X75285; CAAS3040.1; -.
DR EMBL; AF135253; AAD34456.1; -.
DR EMBL; AF135253; AAD34456.1; JOINED.
DR EMBL; AF135249; AAD34456.1; JOINED.
DR EMBL; AF135240; AAD34456.1; JOINED.
DR EMBL; AF135241; AAD34456.1; JOINED.
DR EMBL; AF135242; AAD34456.1; JOINED.
DR EMBL; AF135243; AAD34456.1; JOINED.
DR EMBL; AF135244; AAD34456.1; JOINED.
DR EMBL; AF135245; AAD34456.1; JOINED.
DR EMBL; AF135246; AAD34456.1; JOINED.
DR EMBL; AF135247; AAD34456.1; JOINED.
DR EMBL; AF135248; AAD34456.1; JOINED.
DR EMBL; AF135249; AAD34456.1; JOINED.
DR EMBL; AF135250; AAD34456.1; JOINED.
DR EMBL; AF135251; AAD34456.1; JOINED.
DR EMBL; AF135252; AAD34456.1; JOINED.
DR PIR; A49457; A49457.
DR HSSP; P07207; A49457.
DR MGD; MGI:95488; Fbln2.
DR InterPro; IPR000020; Anaphylatoxin.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF01821; ANATO; 2.
DR Pfam; PF00008; EGF; 6.
DR SMART; SM00104; ANATO; 3.
DR SMART; SM00179; EGF_CA; 9.
DR SMART; SM00001; EGF_like; 2.

```

DR	PROSITE; PS00010;	ASX_HYDROXYL_5.
DR	PROSITE; PS01177;	ANAPHYLATOXIN_1; 3.
DR	PROSITE; PS01178;	ANAPHYLATOXIN_2; 3.
DR	PROSITE; PS00022;	EGF-1; FALSE_NEG.
DR	PROSITE; PS01186;	EGF-2; 5.
DR	PROSITE; PS01187;	EGF_CA; 10.
KM	Signal; Glycoprotein; Extracellular matrix; Plasma; EGF-like domain; Calcium-binding; Alternative splicing; Repeat.	
FT	SIGNAL	1 26
FT	CHAIN	27 1221
FT	DOMAIN	27 434
FT	DOMAIN	27 176
FT	DOMAIN	177 434
FT	DOMAIN	435 477
FT	DOMAIN	478 510
FT	DOMAIN	511 543
FT	DOMAIN	594 635
FT	DOMAIN	669 705
FT	DOMAIN	709 758
FT	DOMAIN	756 800
FT	DOMAIN	801 846
FT	DOMAIN	847 894
FT	DOMAIN	895 937
FT	DOMAIN	938 979
FT	DOMAIN	980 1061
FT	DOMAIN	1019 1061
FT	DOMAIN	1062 1106
FT	DOMAIN	1111 1221
FT	SITE	421 423
FT	DISUPEID	433 462
FT	DISUPEID	436 469
FT	DISUPEID	449 470
FT	DISUPEID	479 508
FT	DISUPEID	492 509
FT	DISUPEID	511 535
FT	DISUPEID	512 542
FT	DISUPEID	525 543
FT	DISUPEID	598 610
FT	DISUPEID	606 619
FT	DISUPEID	621 634
FT	DISUPEID	673 683
FT	DISUPEID	679 692
FT	DISUPEID	694 707
FT	DISUPEID	713 726
FT	DISUPEID	720 735
FT	DISUPEID	742 754
FT	DISUPEID	805 818
FT	DISUPEID	812 827
FT	DISUPEID	833 845
FT	DISUPEID	899 912
FT	DISUPEID	906 921
FT	DISUPEID	923 936
FT	DISUPEID	942 954
FT	DISUPEID	950 963
FT	DISUPEID	965 978
FT	DISUPEID	984 993
FT	DISUPEID	989 1002
FT	DISUPEID	1004 1017
FT	DISUPEID	1023 1035
FT	DISUPEID	1031 1044
FT	DISUPEID	1046 1060
FT	DISUPEID	1066 1079
FT	DISUPEID	1073 1088
FT	DISUPEID	1093 1105
FT	CARBOHYD	179 179
FT	CARBOHYD	497 497
FT	CARBOHYD	737 737
FT	VARSPLIC	1072 1072
FT	CONFLICT	140 159
FT	CONFLICT	348 348
FT	CONFLICT	507 507





```

FT DISULFID 634 645 .BY SIMILARITY.
FT DISULFID 640 654 .BY SIMILARITY.
FT DISULFID 657 669 .BY SIMILARITY.
FT DISULFID 675 686 .BY SIMILARITY.
FT DISULFID 681 695 .BY SIMILARITY.
FT DISULFID 698 709 .BY SIMILARITY.
FT DISULFID 715 726 .BY SIMILARITY.
FT DISULFID 721 735 .BY SIMILARITY.
FT DISULFID 737 750 .BY SIMILARITY.
FT DISULFID 756 767 .BY SIMILARITY.
FT DISULFID 762 776 .BY SIMILARITY.
FT DISULFID 778 791 .BY SIMILARITY.
FT DISULFID 797 808 .BY SIMILARITY.
FT DISULFID 803 817 .BY SIMILARITY.
FT DISULFID 819 832 .BY SIMILARITY.
FT DISULFID 838 850 .BY SIMILARITY.
FT DISULFID 845 859 .BY SIMILARITY.
FT DISULFID 861 873 .BY SIMILARITY.
FT DISULFID 879 891 .BY SIMILARITY.
FT DISULFID 885 900 .BY SIMILARITY.
FT DISULFID 902 915 .BY SIMILARITY.
FT DISULFID 921 933 .BY SIMILARITY.
FT DISULFID 927 942 .BY SIMILARITY.
FT DISULFID 944 957 .BY SIMILARITY.
FT DISULFID 963 975 .BY SIMILARITY.
FT DISULFID 970 984 .BY SIMILARITY.
FT DISULFID 986 1000 .BY SIMILARITY.
FT DISULFID 1101 1114 .BY SIMILARITY.
FT DISULFID 1109 1123 .BY SIMILARITY.
FT DISULFID 1125 1138 .BY SIMILARITY.
FT DISULFID 1144 1155 .BY SIMILARITY.
FT DISULFID 1150 1164 .BY SIMILARITY.
FT DISULFID 1166 1179 .BY SIMILARITY.
FT DISULFID 1298 1309 .BY SIMILARITY.
FT DISULFID 1304 1318 .BY SIMILARITY.
FT DISULFID 1320 1333 .BY SIMILARITY.
FT DISULFID 1339 1354 .BY SIMILARITY.
FT DISULFID 1349 1363 .BY SIMILARITY.
FT DISULFID 1365 1378 .BY SIMILARITY.
FT DISULFID 1365 1378 .BY SIMILARITY.
FT MOD_RES 647 647 .HYDROXYLATION.
FT MOD_RES 810 810 .HYDROXYLATION.
FT CARBOHYD 21 21 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 52 52 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 294 294 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 870 870 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 923 923 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1039 1039 .N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1394 AA; 152791 MW; DFCAR1A40B2C7D1 CRC64;

Query Match 12.08; Score 240.5; DB 1; Length 1394;
Best Local Similarity 27.1%; Pred. No. 1.5e-10;
Matches 85; Conservative 22; Mismatches 96; Indels 111; Gaps 19;

```

```

Db 831 -----HCEIDNECTEDKSVQOR--GDCINTAGSYDCTCPDGFOLDNKTQCQDNECEHP 882
QY 328 DACYPAAEAETEG 341
Db 883 GLCGPGCECLNTEG 896

RESULT 16
PRTS_HUMAN
ID PRTS_HUMAN STANDARD; PRT; 676 AA.
AC P07225; Q15518;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Vitamin K-dependent protein S precursor.
GN PROS1 OR PROS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91084444; Pubmed=2148110;
RA Schmidt D.K., Tatro A.V., Phelps L.G., Tomczak J.A., Long G.L.;
RT "Organization of the human protein S genes.";
RL Biochemistry 29:7845-7852(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91084445; Pubmed=2148111;
RA Ploos van Amstel H.K., Reitsma P.H., der Logt C.P., Bertina R.M.;
RT "Intron-exon organization of the active human protein S gene PS alpha
and its pseudogene PS beta: duplication and silencing during primate
evolution.";
RL Biochemistry 29:7853-7861(1990).
RN [4]
RP SEQUENCE OF 1-42 FROM N.A.
RX MEDLINE=88005138; Pubmed=2820795;
RA Ploos van Amstel H.K., van der Zanden A.L., Reitsma P.H.,
RA Bertina R.M.;
RT "Human protein S cDNA encodes Phe-16 and Tyr 222 in consensus
sequences for the post-translational processing.";
RL FEBS Lett. 222:186-190(1987).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092407; Pubmed=3467362;
RA Hoskins J., Norman D.K., Beckmann R.J., Long G.L.;
RT "Cloning and characterization of human liver cDNA encoding a protein
S precursor.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:349-353(1987).
RN [6]
RP SEQUENCE OF 27-676 FROM N.A.
RX MEDLINE=86313649; Pubmed=2944113;
RA Lundwall A., Deckowski W., Cohen E., Shaffer M., Mahr A., Dahlback B.,
RA Stenflo J., Wydro R.;
RT "Isolation and sequence of the cDNA for human protein S, a regulator
of blood coagulation.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6716-6720(1986).
RN [7]
RP VARIANT HEERLEN.
RX MEDLINE=90335440; Pubmed=2143091;
RA Bertina R.M., Ploos van Amstel H.K., van Wijngaarden A.,
RA Coenen J., Leemhuis M.P., Deutz-Ferlouw P.P., van der Linden I.K.,
RA Reitsma P.H.;
RT "Heerlen polymorphism of protein S, an immunologic polymorphism due
to dimorphism of residue 460.";
RL Blood 76:538-548(1990).
RN [8]

```





QY 71 SEIRLELEIGLCESSDFECNOMLE--AOEHLBAMWL-----OLKSEYP 113  
 DB 57 -----ECIEBELCNKE--EAREVEFENDEPETYFYFKYLVCLARSPTGTLFTAROSTNAYP 108  
 QY 114 DLFEMFCVKTLKVCSSPGTYGPDCLACGGSGRSCNGHCS-GDGSFQDGGSCRCHEWY 172  
 DB 109 DLRS/-CVNAIPDCCSP-----LPCNEDGYMSCKKG--KASFTCTCKRGW 149  
 QY 173 QGPIG----TDCMDGYFSLNETHSICTACDESKTCSGLTNDCCGCEYGVWL--DEG 226  
 DB 150 QGKCEPDINCKD-----PSNINGGCSQICDNTPG--SYHC--SCKNGFVWLNMK 197  
 QY 227 ACVDVDECAAPPCSAOAFCKNANGSYTCECDSSCVGCTGEGRGNCKECISGYA--RE 284  
 DB 198 DCKVDKSLKPSICGTA-VCKNIPGDEPC-----ECPEGTRYMLK 237  
 QY 285 HGCCADVDECSLAECTYCRKNENCTNPGSYVVC--PDGEETED-----ACVP 332  
 DB 238 SKCEDIDECSS--ENMCA---QLCVNYPGGYCYCDGKRGKLAQDOKSCCEVVSCLP 290

RESULT 17  
 PK5\_RAT  
 ID PK5\_RAT STANDARD: PRT: 1877 AA.  
 AC P41413; 062914;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
 DE (Proprotein convertase PC5) (Subtilisin/kexin-like protease PC5)  
 DE (Convertase PC5) (rPC5) (PC6) (Fragments).  
 GN PCSK5.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCBI\_TaxID=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND TISSUE-SPECIFICITY.  
 RC TISSUE=Adrenal gland;  
 RX MEDLINE=93342056; PubMed=8341687;  
 RA Luson J., Vleau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
 RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a  
 RT candidate proprotein convertase expressed in endocrine and  
 RT nonendocrine cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
 RN [2]  
 RP SEQUENCE OF 1676-1877 FROM N.A. (LONG ISOFORM).  
 RC TISSUE=Adrenal gland;  
 RA De Bie I., Marcinkiewicz M., Nakayama K., Lazure C., Seidah N.G.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=20214819; PubMed=10749928;  
 RA Xiang Y., Molloy S.S., Thomas L., Thomas G.;  
 RT "The PC6b cytoplasmic domain contains two acidic clusters that direct  
 RT sorting to distinct trans-Golgi network/endosomal compartments.";  
 RL Mol. Biol. Cell 11:1257-1273(2000).  
 RN [4]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE=97166043; PubMed=9013936;  
 RA Zheng M., Seidah N.G., Platar J.E.;  
 RT "The developmental expression in the rat CNS and peripheral tissues of  
 RT proteases PCS and PACFA mRNAs: comparison with other proprotein  
 RT processing enzymes.";  
 RL Dev. Biol. 181:268-283(1997).  
 CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY  
 CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE  
 CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE  
 CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED  
 CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION  
 CC OF GROWTH FACTORS.  
 CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR  
 CC PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS, WHERE XAA

CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.  
 CC -1- SUBCELLULAR LOCATION: PCS5 IS SECRETED THROUGH THE REGULATED  
 CC SECRETORY PATHWAY. PCS5 IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO  
 CC A PERINUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH  
 CC EARLY ENDOSOMES (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PCS5B/LONG (SHOWN HERE)  
 CC AND PCS5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE INTESTINE, BRAIN, ADRENAL  
 CC GLAND, ANTERIOR PITUITARY, THYROID, OVARIES, TESTIS AND LUNG.  
 CC HIGHEST LEVELS ARE FOUND IN THE GUT, DUODENUM, JEJUNUM AND ILEUM.  
 CC EXPRESSION IS HIGHER IN FEMALE THAN IN MALE REPRODUCTIVE ORGANS.  
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT E9 IN HIGHLY RESTRICTED  
 CC REGIONS OF THE NEURAL TUBE, IN CAUDAL MYOTOMES, AND AT THE  
 CC MATERNO-EMBRYONIC JUNCTION OF THE UTERUS. AT E10, RESTRICTED  
 CC EXPRESSION IS DETECTED IN THE OPTIC AND OTIC VESTICLES, THE ROOF OF  
 CC MIDBRAIN, AND TRUNK MYOTOMES. BY MIDGESTATION (E13-E16),  
 CC EXPRESSION IN THE DEVELOPING NERVOUS SYSTEM HAS EXPANDED TO  
 CC MULTIPLE REGIONS INCLUDING HIPPOCAMPUS, THALAMUS, HYPOTHALAMUS,  
 CC BRAIN STEM, AND SPINAL CORD. EXPRESSION IS ALSO DETECTED IN  
 CC SEVERAL PERIPHERAL ORGAN SYSTEMS, INCLUDING GUT, LUNG, ADRENAL AND  
 CC KIDNEY PRIMORDIA.  
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE  
 CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC  
 CC RETICULUM.  
 CC -1- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN  
 CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS  
 CC WITH THE TGN SORTING PROTEIN PACS-1.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
 CC SUBTILASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation/  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; L14933; AAA99906.1; -;  
 CC EMBL; U47014; AAA87888.1; -;  
 CC PIR; B48225; B48225.  
 CC HSSP; Q99405; IMPT.  
 CC MEROPS; S08.076; -;  
 CC InterPro; IPR002174; Eurin-like.  
 CC InterPro; IPR002884; P-domain.  
 CC InterPro; IPR000209; Peptidase\_S8.  
 CC Pfam; PF01483; P; 1.  
 CC Pfam; PF00082; Peptidase\_S8; 1.  
 CC PRINTS; PR00723; SUBTILISIN.  
 CC PRODOM; PD000717; P-domain; 1.  
 CC SMART; SM00261; FU; 6.  
 CC PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 CC PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 CC PROSITE; PS00138; SUBTILASE\_SER; 1.  
 CC KW Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal;  
 CC Cleavage on pair of basic residues; Alternative splicing; Repeat;  
 CC Transmembrane.  
 CC KW SIGNAL .1 34  
 CC PROPEP 35 116  
 CC CHAIN 117 1877  
 CC FT  
 CC DOMAIN 117 1768  
 CC TRANSMEM 1769 1789  
 CC DOMAIN 1790 1877  
 CC DOMAIN 117 452  
 CC DOMAIN 464 602  
 CC DOMAIN 638 1753  
 CC DOMAIN 1825 1844  
 CC DOMAIN 1856 1877  
 CC SITE 116 117  
 CC SITE 521 523  
 CC ACT\_SITE 173 173  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC PROPROTEIN CONVERTASE SUBTILISIN/KEXIN  
 CC TYPE 5.  
 CC EXTRACELLULAR (POTENTIAL).  
 CC POTENTIAL.  
 CC CYTOPLASMIC (POTENTIAL).  
 CC CATALYTIC.  
 CC CY5-RICH MOTIF (CRM) REGION.  
 CC AC 1.  
 CC AC 2.  
 CC CLEAVAGE (AUTO-) (BY SIMILARITY).  
 CC CELL ATTACHMENT SITE (POTENTIAL).  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).

```

FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 388 388 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 854 854 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1710 1710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1732 1732 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 878 915 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX ->
FT VARSPLIC 915 915 ATEESAEAGFCMLYKNNKLCQRKVLQQLCCKTCTGQS
FT VARSPLIC 916 1877 (IN ISOFORM PCSA).
FT VARSPLIC 1877 AA: 207888 MW: 890955DC60534444 CRC64:
SQ SEQUENCE

Query Match 11.8%; Score 236; DB 1: Length 1877;
Best Local Similarity 28.8%; Pred. No. 2.5e-10;
Matches 66; Conservative 16; Mismatches 79; Indels 70; Gaps 12;

QY 125 KVC---CSPCTYGPDLACGSGRPSGNGHSGSGDSRGDSCRCRHMVYQGPPLCTDCM 181
DB 670 RIVSSCPPEHFHADKKRC-----KCAPN-----CESCFESHADQCLSK 710
QY 182 DGYFSSLRNETHS-----ICTACDSCKTSGLTNRNDCGECEVWVLE 225
DB 711 YGYF--LNEETSSCAVACPEGYODIKKIKGKSEKCTCTGFHN--CTECKGGLSLG 766
QY 226 GACVDVDECAEPFPCSAOFCNANGSYCEGCDSSCVCTGEGONCKECLSGAREH 285
DB 767 SRC-----SVTCEDQFF--SGHCGPCRHRCATCAGADGCTNCTEGYMER 813
QY 286 GGCADYDECSIA-----EKTGVKNENY--NTPGSYVC--VCPDGF 323
DB 814 GRC--VQSGCVSYLIDHSLGEGYKSKCRCDNCLTNGFPKNCSCPSQY 862

RESULT 18
ID PCSK5_HUMAN STANDARD; PRT; 913 AA.
AC 092824: Q13527:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Proteinase PC5) (Subtilisin/kexin-like protease PC5)
DE (Convertase PC5) (PC6) (hPC6).
GN PCSK5 OR PC5 OR PC6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RX MEDLINE=96353880; PubMed=8755538;
RA Miranda L., Wolf J., Pichuanes S., Duke R., Franzusoff A.;
RT "Isolation of the human PC6 gene encoding the putative host protease
RT for HIV-1 gp160 processing in CD4+ T lymphocytes."
RL Proc. Natl. Acad. Sci. U.S.A. 93:7695-7700(1996).
RN [2]
RP REVISIONS.
RA Franzusoff A., Miranda L., Wolf J., Pichuanes S., Lu Y., Duke R.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 15-913 FROM N.A.
RA Reudelhuber T.L.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR

```

```

CC PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-2AA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
CC -!- SUBCELLULAR LOCATION: Secreted (by similarity).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS SEEMS TO BE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN T-LYMPHOCYTES.
CC -!- DOMAIN: THE PROPEPTIDASE DOMAIN ACTS AS AN INTRAOLICULAR CHAPERONE
CC ASSISTING IN THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION-
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sdb.ch/announce/
CC OR SEND AN EMAIL TO license@isb-sdb.ch).
CC -----
DR EMBL: U56387; AAC50643.2; -
DR EMBL: U49114; AAA91807.1; -
DR HSP: Q99405; IMP1.
DR MIM: 600488; -
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002884; P_domain.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF01483; P: 1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR ProDom: PD000717; P_domain; 1.
DR SMART: SM00261; FU; 5.
DR PROSITE: PS00136; SUBTILASE ASP; 1.
DR PROSITE: PS00137; SUBTILASE HIS; 1.
DR PROSITE: PS00138; SUBTILASE SER; 1.
KW Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Cleavage on pair of basic residues; Repeat.
FT SIGNAL 1 32
FT PROPEP 33 114
FT CHAIN 115 913
FT TYPE 5.
FT CATALYTIC.
FT DOMAIN 115 454
FT DOMAIN 462 600
FT SITE 636 913
FT SITE 114 115
FT SITE 519 521
FT ACT_SITE 171 171
FT ACT_SITE 212 212
FT ACT_SITE 212 212
FT ACT_SITE 386 386
FT CARBOHYD 225 225
FT CARBOHYD 381 381
FT CARBOHYD 665 665
FT CARBOHYD 752 752
FT CARBOHYD 802 802
FT CARBOHYD 852 852
FT CONFLICT 118 118
FT CONFLICT 121 121
FT CONFLICT 511 511
FT CONFLICT 601 601
SQ SEQUENCE 913 AA: 101775 MW: 21389264CAD7546C CRC64:

Query Match 11.8%; Score 236; DB 1: Length 913;
Best Local Similarity 25.8%; Pred. No. 2.1e-10;
Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;

QY 38 VDKFNGWVDTAKKNGGNTAMEEKLTKYSESLRLLELGLCESSDFPCNDMLEAQ 97
DB 612 VERFRRSVREDDPDYDGTEDYAGP-----CDPECSV-----GCDGPPDHCHNDCLH-- 658
QY 98 EHLLEWMLQLMSFYDLDFEWFCVKTLL-----KVC--CSP-----GTYGPCCLAC 140

```

Db 659 -----YYLKLNN-----TRICVSSCPGHHYHAKKRCRCAPNCSCSCFSGHGDQMSC 707  
 Qy 141 QGQ-----SQRPCSGNGHSCGSDGSRG--DGSRC-----CHMOYQPLCTDCMDGYESSLRN 190  
 Db 708 KYGFLNEEINSCVTHCP-DGSIYDTKKNLCRCSEKCTCTEHNTECHDGL--SLQG 764  
 Qy 191 ETHSI-----CTACDESKCTKSGLTNRDCECEVWVLDBGACVYDECAAP 238  
 Db 765 SRCVSCDEGRYFNGDCQPCGHRFCATCAGAGADGCINCTGTYEMEDGRVQ----- 816  
 Qy 239 PCGAAPFCNA--NGSYTEECDDSCVCGTGECPGCKEISYAEHGO-----CAD 290  
 Db 817 -SCSIYFDPHSENGSKCKCDISCTCNGPGFKNCTSCPSGYLLDLGCMGAIKCD 875  
 Qy 291 VDECSLAEK---TCVRKNENYNTPGSYVC 317  
 Db 876 ATEESMAEGGFCMLYKKNKNCQKRVLDQIC 905

RESULT 19  
 PRS\_MOUSE STANDARD; PRT: 675 AA.  
 ID PRS\_MOUSE  
 AC Q08761; P43483;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Vitamin K-dependent protein S precursor.  
 GN PROS1 OR PROS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94198297; PubMed=8148380;  
 RA Chu M.D., Sun J., Bird P.I.;  
 RT "Cloning and sequencing of a cDNA encoding the murine vitamin K-  
 RT dependent protein S.";  
 RL Biochim. Biophys. Acta 1217:325-328(1994).  
 RN [2]  
 RP SEQUENCE OF 33-675 FROM N.A.  
 RX MEDLINE=94302659; PubMed=8029814;  
 RA La D., Schmidel D.K., Long G.L.;  
 RT "Structure of mouse protein S as determined by PCR amplification and  
 RT DNA sequencing of cDNA.";  
 RL Thromb. Res. Suppl. 74:135-142(1994).  
 CC -!- FUNCTION: ANTICOAGULANT PLASMA PROTEIN. IT IS A COFACTOR TO  
 CC ACTIVATED PROTEIN C IN THE DEGRADATION OF COAGULATION FACTORS VA  
 CC AND VIIIA. IT HELP TO PREVENT COAGULATION AND STIMULATING  
 CC FIBRINOLYSIS.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- TISSUE SPECIFICITY: PLASMA.  
 CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: Z25469; CAA80961.1; -;  
 DR EMBL: L27439; AAA40006.1; -;  
 DR HSSP: P00740; ICFH.  
 DR MGD: MGI:1095733; Pros1.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR002383; GLA\_blood.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR000294; VitK\_dep\_GLA.

DR Pfam: PF00008; EGF; 4.  
 DR Pfam: PF00594; gla; 1.  
 DR Pfam: PF00054; laminin\_G; 1.  
 DR PRINTS: PR00001; GLABLOOD.  
 DR SMART: SM00179; EGF\_CA; 3.  
 DR SMART: SM00001; EGF\_like; 1.  
 DR SMART: SM00069; GLA; 1.  
 DR SMART: SM00282; Lamg; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 4.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; 3.  
 DR PROSITE: PS01187; EGF\_CA; 3.  
 DR PROSITE: PS00011; GLU\_CARBOXYLATION; 1.  
 DR PROSITE: PS00025; LAM\_G\_DOMAIN; 2.  
 DR Plasma: Gamma-carboxylutamic acid: Calcium; Vitamin K; Zymogen;  
 KW Blood coagulation; Hydroxylation; Glycoprotein; Signal; Repeat;  
 KW EGF-like domain.  
 FT SIGNAL 1 24  
 FT PROPEP 25 41  
 FT CHAIN 42 675  
 FT DOMAIN 42 86  
 FT DOMAIN 87 116  
 FT DOMAIN 117 135  
 FT DOMAIN 157 200  
 FT DOMAIN 201 242  
 FT DOMAIN 243 283  
 FT DOMAIN 299 475  
 FT DOMAIN 484 665  
 FT MOD\_RES 47 47  
 FT MOD\_RES 48 48  
 FT MOD\_RES 55 55  
 FT MOD\_RES 57 57  
 FT MOD\_RES 60 60  
 FT MOD\_RES 61 61  
 FT MOD\_RES 66 66  
 FT MOD\_RES 67 67  
 FT MOD\_RES 70 70  
 FT MOD\_RES 73 73  
 FT MOD\_RES 77 77  
 FT MOD\_RES 136 136  
 FT DISULEID 121 134  
 FT DISULEID 126 143  
 FT DISULEID 145 154  
 FT DISULEID 161 175  
 FT DISULEID 171 184  
 FT DISULEID 186 199  
 FT DISULEID 205 217  
 FT DISULEID 212 226  
 FT DISULEID 228 241  
 FT DISULEID 247 256  
 FT DISULEID 252 265  
 FT DISULEID 267 282  
 FT CARBOHYD 499 499  
 FT CARBOHYD 509 509  
 FT CONFLICT 493 493  
 FT SEQUENCE 675 AA; 74934 MM; 79D51203E85AF31F CRC64;

Query Match 11.68; Score 232; DB 1; Length 675;  
 Best Local Similarity 26.38; Pred. No. 3; de-10;  
 Matches 91; Conservative 30; Mismatches 119; Indels 106; Gaps 20;

OY 11 LRLPLLLPAPPAKPPCHRCGLVDKFNQGVNVTAKKNGGNTAMEKTLKYES 70  
 DB 11 LAACIALATVPSEPT-----NFLSKERASVLYVKRR---ANTLFEEETMKNLER 56  
 OY 71 SEIRLLEIEGLCESSDFECNOMLEQEEH-----LEKMWL-----QLKSEFP 113  
 DB 57 -----ECIEELCKNE--FAREVENNPETDYFPYKYLCLGAFRGSFHARQSANAP 108  
 OY 114 DLFEMFCVTLKVCSPGYGPD-CLACOGSQRPCSGNHCSGDSDRGDSRCRHMVY 172  
 DB 109 DLRS--CVMAISDQCPICNEGYLACODG-----QAATFCCKPQM 149  
 OY 173 QGPLC---TDCMDGYFSSLRNETHSICPACDESCCTGSLNRDCEGEVGVTL--DEG 226  
 DB 150 QGDRQCYDNECKD-----PSNVNGGCSQICDNTPG--SYHC--SCKRGFAMLPNKK 197  
 OY 227 ACYVDVDECAEPPPCSAAPCKRANANGSYCEEDSSVCCTGEGPNCCECISGYARENG 286  
 DB 198 DCKDDECALKPSVCGTA-VCKNIPGDFEC-----ECPDGYRYDPS 237  
 OY 287 --QCADVDECSLAECTVCKRKNENCYTPGSYVCV--PDGFEETED 328  
 DB 238 SKSKRVDECS--ENMKA---QLCVAFPGGYSCYCDKKGFKLAQD 278

RESULT 20  
 ID NID2\_HUMAN STANDARD: PRT: 1375 AA.  
 AC 014112: 043710:  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nidogen-2 precursor (NID-2) (Osteonidogen).  
 GN NID2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.  
 RX MEDLINE=98406162; PubMed=9733643;  
 RA Kohfeldt E., Sasaki T., Goehring W., Timpi R.;  
 RT "Nidogen-2: a new basement membrane protein with diverse binding  
 RT properties.";  
 RL J. Mol. Biol. 282:99-109(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=cartilago bone;  
 RA Ohno I., Hashimoto J., Yakaoka K., Ochi T., Okubo K., Matsubara K.;  
 RT "The cloning and characterization of a cDNA for the novel bone matrix  
 RT protein: osteonidogen.";  
 RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ohno I., Okubo K., Matsubara K.;  
 RT "Human osteonidogen gene: Intron-exon junctions and chromosomal  
 RT localization.";  
 RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: CELL ADHESION GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED  
 CC IN BASEMENT MEMBRANES. BINDS TO COLLAGENS I AND IV, TO PERLECAN  
 CC AND TO LAMININ 1. DOES NOT BIND FIBRILIN. IT PROBABLY HAS A ROLE  
 CC IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.  
 CC -1- TISSUE SPECIFICITY: HEART, PLACENTA AND BONE. LESS IN PANCREAS,  
 CC KIDNEY AND SKELETAL MUSCLE.  
 CC -1- PTM: HIGHLY N- AND O-GLYCOSYLATED.  
 CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-I DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ223500; CAA11418.1; -  
 DR EMBL: D64425; BAA13087.1; -  
 DR EMBL: AB009779; BAA24112.1; JOINED.  
 DR EMBL: AB009778; BAA24112.1; JOINED.  
 DR EMBL: AB009779; BAA24112.1; JOINED.  
 DR EMBL: AB009780; BAA24112.1; JOINED.  
 DR EMBL: AB009781; BAA24112.1; JOINED.  
 DR EMBL: AB009782; BAA24112.1; JOINED.  
 DR EMBL: AB009783; BAA24112.1; JOINED.  
 DR EMBL: AB009784; BAA24112.1; JOINED.  
 DR EMBL: AB009785; BAA24112.1; JOINED.  
 DR EMBL: AB009786; BAA24112.1; JOINED.  
 DR EMBL: AB009787; BAA24112.1; JOINED.  
 DR EMBL: AB009788; BAA24112.1; JOINED.  
 DR EMBL: AB009789; BAA24112.1; JOINED.  
 DR EMBL: AB009790; BAA24112.1; JOINED.  
 DR EMBL: AB009791; BAA24112.1; JOINED.  
 DR EMBL: AB009792; BAA24112.1; JOINED.  
 DR EMBL: AB009793; BAA24112.1; JOINED.  
 DR EMBL: AB009794; BAA24112.1; JOINED.  
 DR EMBL: AB009795; BAA24112.1; JOINED.  
 DR EMBL: AB009796; BAA24112.1; JOINED.  
 DR EMBL: AB009797; BAA24112.1; JOINED.  
 DR EMBL: AB009798; BAA24112.1; JOINED.  
 DR HSP: P35555; IEMN.  
 DR MIM: 605399; -  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-like.  
 DR InterPro: IPR000033; IgL\_receptor\_rep.  
 DR InterPro: IPR003886; Nidogen\_ext.  
 DR InterPro: IPR00716; Thyroglobulin\_1.  
 DR Pfam: PF00008; EGF; 5.  
 DR Pfam: PF00058; IgL\_recept\_b; 4.  
 DR Pfam: PF00086; thyroglobulin\_1; 2.  
 DR SMART: SM00179; EGF\_CA; 2.  
 DR SMART: SM00001; EGF\_like; 3.  
 DR SMART: SM00135; LY; 4.  
 DR SMART: SM00539; NIDO; 1.  
 DR SMART: SM00211; TY; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1; 2.  
 DR PROSITE: PS01186; EGF\_2; 4.  
 DR PROSITE: PS01187; EGF\_CA; 2.  
 KW Basement membrane; Extracellular matrix; Glycoprotein; Signal;  
 KW Calcium-binding; Repeat; EGF-like domain; Cell adhesion.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1375 NIDOGEN-2.  
 FT DOMAIN 484 524 EGF-LIKE 1.  
 FT DOMAIN 759 800 EGF-LIKE 2.  
 FT DOMAIN 801 843 EGF-LIKE 3.  
 FT DOMAIN 848 891 EGF-LIKE 4.  
 FT DOMAIN 932 930 EGF-LIKE 5.  
 FT DOMAIN 936 1006 CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 1015 1085 THYROGLOBULIN TYPE I 1.  
 FT DOMAIN 1157 1200 THYROGLOBULIN TYPE I 2.  
 FT DOMAIN 1201 1243 LDL-RECEPTOR YWTD MOTIF 1.  
 FT DOMAIN 1244 1288 LDL-RECEPTOR YWTD MOTIF 2.  
 FT DOMAIN 1289 1328 LDL-RECEPTOR YWTD MOTIF 3.  
 FT DOMAIN 1329 1373 LDL-RECEPTOR YWTD MOTIF 4.  
 FT DOMAIN 1373 1776 LDL-RECEPTOR YWTD MOTIF 5.  
 FT DISULFID 763 776 BY SIMILARITY.  
 FT DISULFID 770 786 BY SIMILARITY.  
 FT DISULFID 788 799 BY SIMILARITY.  
 FT DISULFID 805 818 BY SIMILARITY.  
 FT DISULFID 812 827 BY SIMILARITY.  
 FT DISULFID 829 842 BY SIMILARITY.  
 FT DISULFID 852 867 BY SIMILARITY.  
 FT DISULFID 859 877 BY SIMILARITY.  
 FT DISULFID 879 890 BY SIMILARITY.





```

Db 1266 TCSQCVSRNCSNCKGLELQNGECRTTCADGYSD-----RGICAKCYLSCHTCSGPRR 1320
Qy 212 RDGECEVGVNLDEGACVDDECAEPPPCSAAPFCNANSGYCEBDCSSVCVCTEGP 271
Db 1321 NOCVQCPAGWQLAAGEC-----HPECPEGFY-----KSDFCQCHYCKTKCNDAGR 1367
Qy 272 GNCKECTISGVAHEGQCAD-----VDECSLAERTCVKRNKNCYNTPGSYCVCPDGFEE 325
Db 1368 LACTSCPHMLDGLCLMCECLSSGYDTTSATCTCHDCSCRCG-PGQFSC----- 1418
Qy 326 TEDACVPP 333
Db 1419 --KGCVPP 1424

RESULT 23
NTCL_MOUSE
ID NTCL_MOUSE STANDARD; PRT; 2531 AA.
AC 001705;
DT 01-NOV-1996 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (NOTCH protein).
GN NOTCH1 OR NOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxId=10090;
RX SEQUENCE FROM N.A.
RP TISSUE=Embryo;
RX MEDLINE=93194170; PubMed=8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RL homolog of Drosophila Notch."
RN Genomics 15:259-264(1993).
RP SEQUENCE OF 1551-2170 FROM N.A.
RX TISSUE=Embryo;
RX MEDLINE=93048835; PubMed=1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development."
RL Development 115:737-744(1992).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z11886; CAAT7941.1; -.
DR HSSP: P00740; IEDM.
DR MGD: MG1:97363; Notch1.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR001438; EGF_TI.
DR InterPro: IPR000800; Notch.

DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 35.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00010; EGFBL00.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 3.
DR SMART: SM00179; EGF_CA; 23.
DR SMART: SM00001; EGF_Like; 11.
DR SMART: SM00004; NL; 2.
DR PROSITE: PSS0088; ANK_REPEAT; 2.
DR PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 22.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS01186; EGF_2; 27.
DR PROSITE: PS01187; EGF_CA; 21.
DR Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 2531
FT DOMAIN 19 1725
FT TRANSFEM 1726 1746
FT DOMAIN 1747 2531
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 255
FT DOMAIN 257 293
FT DOMAIN 295 333
FT DOMAIN 335 371
FT DOMAIN 372 410
FT DOMAIN 412 450
FT DOMAIN 452 488
FT DOMAIN 490 526
FT DOMAIN 528 564
FT DOMAIN 566 601
FT DOMAIN 603 639
FT DOMAIN 641 676
FT DOMAIN 678 714
FT DOMAIN 716 751
FT DOMAIN 753 789
FT DOMAIN 791 827
FT DOMAIN 829 867
FT DOMAIN 869 905
FT DOMAIN 907 943
FT DOMAIN 945 981
FT DOMAIN 983 1019
FT DOMAIN 1021 1057
FT DOMAIN 1059 1095
FT DOMAIN 1097 1143
FT DOMAIN 1145 1181
FT DOMAIN 1183 1219
FT DOMAIN 1221 1265
FT DOMAIN 1267 1305
FT DOMAIN 1307 1346
FT DOMAIN 1348 1384
FT DOMAIN 1387 1426
FT DOMAIN 1449 1462
FT DOMAIN 1445 1480
FT REPEAT 1481 1522
FT REPEAT 1523 1562
FT REPEAT 1517 1947
FT REPEAT 1949 1979
FT REPEAT 1983 2012
FT REPEAT 2016 2045
FT REPEAT 2049 2078
FT DISULFID 31 46
FT DISULFID 63 74
FT DISULFID 68 87
FT DISULFID 89 98
FT DISULFID 106 117

```



```

FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 855 866 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 1007 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.

```

```

Query Match 11.3%; Score 226; DB 1; Length 2531;
Best Local Similarity 23.9%; Pred. No. 3.1e-09;
Matches 77; Conservative 21; Mismatches 98; Indels 126; Gaps 17;

OY 126 VC-CSPRGTPGPPCL-----ACGGGQRP--CSGNCHSGDGSNGSGRCHNGYGGPIC- 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 163 ICRPPGFGPPCRQDVNEC---SQNPGLCRHGCHNE--TGSYRCACCATYHGPICE 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 178 -----TDCADGYFSLRNETHSICVA-----CDESKTSGLTNRDCCGEVGM 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 217 LPYVPCSPPCNGMFCRTGTGTHHCACLPGFAGNCEENVDDCPG-----NNCKNG- 269
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 222 VDEGACV-----DVDECAEPPPCSAQFCFNANGSYTC----- 256
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 270 ---GACVDGVNTYNCRPPEVTGQCTEDVDECOLMPNACONAGCTNHTGNGVCVN 325
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 257 ---EECDSSCVGCGEGS---PGNCKE-----CISGYAREHG 286
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 326 GWTEDCSENIDDCASACFOGATCHDRVASFCECPHGRIGLCLHLKHACISNPCNBGS 385
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 287 QC-----ADVDECSLAKETCYRKNNENYNTPGSYVCYCPDGF 323
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 386 NCDTNPVNGKRICICPSGYTGACSDVDVDECDLGANRCHNACK-CLNTLGSPFCGCLGY 444
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 324 -----EETEDACVP-PAEAAT 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 445 TGPCEIDVNECISNPCONDAT 466
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 24

PRTS\_RABIT STANDARD; PRT; 646 AA.

AC P98118; 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE 01-MAR-2002 (Rel. 41, Last annotation update)

PR Vitamin K-dependent protein S precursor (Fragment).

GN PROS1 OR PROS.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI\_Taxid=9986;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=94039141; PubMed=8223642;

HE He X., Dahlback B.;

RT "Molecular cloning, expression and functional characterization of

rabbit anticoagulant vitamin K-dependent protein S.;"

RL Eur. J. Biochem. 217:857-865(1993).

CC -!- FUNCTION: ANTI- COAGULANT PLASMA PROTEIN: IT IS A COFACTOR TO

ACTIVATED PROTEIN C IN THE DEGRADATION OF COAGULATION FACTORS VA

AND VIIIA. IT HELP TO PREVENT COAGULATION AND STIMULATING

FIBRINOLYSIS.

CC -!- FUNCTION: PROTEIN S INTERACTS WITH C4B-BINDING PROTEIN, A

REGULATOR OF THE COMPLEX SYSTEM. IN RABBIT PLASMA HOWEVER, PROTEIN

S APPEARS TO BE PRESENT ONLY IN FREE FORM.

CC -!- SUBCELLULAR LOCATION: Extracellular.

CC -!- TISSUE SPECIFICITY: PLASMA.

CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z26485; CAAB1259.1; ..

CC HSSP: P00740; ICRH.

DR InterPro: IPR000152; Asx\_hydrol.

DR InterPro: IPR000561; EGF-like.



Query Match	11.2%;	Score 225.5;	DB 1;	Length 646;
Best Local Similarity	28.8%;	Pred. No. 9.3e-10;		
Matches	69;	Conservative	24;	Mismatches 72;
				Indels 75;
				Gaps 16

```

Oy 112 YPDLFEMFCVTKLKKCCSPGRTYGPDLACGGGSGRPGSGNGHCS--GDSRQDGSRCRHM 170
Db 78 YPDLRS--CVNAIPDQCNP-----LPCEEGLNCKDG--QATFFCIKRP 118
Oy 171 GYOGFGLC-----TDCKDGGFESSLRNETHSICTAGCDSCKTCGSLTNRDGGCEGVGWL--D 224
Db 119 GWGKCKECLDINECKD-----PTINGGCSOICDNITAG--SYHC-SCKSGFVILAN 166
Oy 225 EGACVADVECAAEPPCASAQFCKNANNGSYTCECDSSCVGCTGEGPNCKECISGVARE 284
Db 167 EKDKCDKMDCEGSKKPSVCGTA--VCKNTPEDDFEC-----ECSGSYRYN 206
Oy 285 --HGCCADVDECSLAERTKVRKNENCYTPGSGYVC--PDGFEETED-----ACVP 332
Db 207 PTAKSCEDIDECSS--ENMCA--QLCVNYPGGYSCYCDGKKGFKLAODKSCSAEAVPCLP 261

RESULT 25
NID2_MOUSE STANDARD; PRT; 1403 AA.
ID NID2_MOUSE
AC 088322;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nidogen-2 precursor (NID-2) (Entactin-2).
GN NID2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98297167; PubMed=9633511;
RA Kimura N., Toyoshima T., Kojima T., Shlman M.;
RT "Entactin-2: a new member of basement membrane protein with high
RL homology to resactin/nidogen."
RT Exp. Cell Res. 241:36-45(1998).
CC -1- FUNCTION: CELL ADHESION GLYCOPROTEIN. MIGHT BE INVOLVED IN
CC OSTEOBLAST DIFFERENTIATION. IT PROBABLY HAS A ROLE IN CELL-
CC EXTRACELLULAR MATRIX INTERACTIONS.
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
CC -1- PTM: HIGHLY N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-I DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation in
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
CC EMBL, AB017202; BAA32609.1; -
CC HSSP; P00736; IAPQ.
DR MGD: MGI:1298229; NID2.
DR InterPro: IPR000152; Asx_Hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR000033; LDL_receptor_rep.
DR InterPro: IPR003886; Nidogen_ext.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF000058; EGF_5.
DR Pfam: PF00058; LDL_recept_b; 3.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR SMART: SM00179; EGF_CA; 2.
DR SMART: SM00001; EGF_Like; 3.
DR SMART: SM00535; LY; 4.
DR SMART: SM00139; NIDO; 1.
DR SMART: SM00211; TY; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; 3.
DR PROSITE: PS01186; EGF_2; 4.

```

DR PROSITE: PS01187; EGF\_CA; 2.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1; 2.  
 KW Basement membrane; Extracellular matrix; Glycoprotein; Signal;  
 KW Calcium-binding; Repeat; EGF-like domain; Cell adhesion.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1403  
 FT DOMAIN 507 547  
 FT DOMAIN 782 823  
 FT DOMAIN 824 866  
 FT DOMAIN 871 914  
 FT DOMAIN 915 953  
 FT DOMAIN 964 1034  
 FT DOMAIN 1043 1113  
 FT DOMAIN 1185 1228  
 FT DOMAIN 1229 1271  
 FT DOMAIN 1272 1316  
 FT DOMAIN 1317 1356  
 FT DOMAIN 1357 1401  
 FT DISULFID 786 799  
 FT DISULFID 793 809  
 FT DISULFID 811 822  
 FT DISULFID 828 841  
 FT DISULFID 835 850  
 FT DISULFID 852 865  
 FT DISULFID 875 890  
 FT DISULFID 882 900  
 FT DISULFID 902 913  
 FT DISULFID 919 930  
 FT DISULFID 924 939  
 FT DISULFID 941 952  
 FT CARBOHYD 681 681  
 FT CARBOHYD 716 716  
 FT CARBOHYD 726 726  
 FT CARBOHYD 1152 1152  
 FT SITE 946 948  
 FT SEQUENCE 1403 AA; 154248 MW; 48f48243905c7e10 CRC64;

Query Match 11.2%; Score 224.5; DB 1; Length 1403;  
 Best Local Similarity 29.0%; Pred. No. 2.3e-09;  
 Matches 70; Conservative 15; Mismatches 67; Indels 89; Gaps 12;

QY 147 PMSGNHCSSGDSRGSGSCRCRMGYQGPLCTDCMDGYSSLSRNETHSICTACDESKTC 206  
 DB 785 PCY-----DGSHTCDTTRARHPG-----TGVDYTC--- 809  
 QY 207 SGLTRNDGCEKGVWLDGACVYDVECAAPPPCSAOGCKNANGSYTCSECDG----- 261  
 DB 810 -----ECTPRGQGDGRSCYDVNVECATGFRRCGPNVCVNLVSYRC-ECRSGYEFA 859  
 QY 262 ----SCV-----GCTGEGPNCCKE-----CISGYAREHGQCADVDEC 294  
 DB 860 DDOHTCILAPPPNCFDSHTCAPREGQARCIHHGSSFSFSCALPFGIFGHCSPVDGC 919  
 QY 295 SLAEKTCVKNKNCYTPPSYVCVCP-----DGFETEDACVPPRAAEATGESPTQLDS 349  
 DB 920 --AENRC-HEAAICVYTPSSFSRCRCOPGVRGDFHCTSDT-VP---EDSISGLKPCYEYO 972  
 QY 350 R 350  
 DB 973 R 973

RESULT 26  
 NOTC\_XENLA STANDARD; PRT; 2524 AA.

AC P21783;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 16-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neurogenic locus notch protein homolog precursor (XORCH protein).  
 GN XORCH.  
 OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_Taxid=8335;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90385285; PubMed=2402639;  
 RA Coffman C., Harris W., Kintner C.;  
 RT "Xorch, the Xenopus homolog of Drosophila notch.";  
 RN Science 249:1438-1441(1990).  
 RP REVISIONS TO 1759-1782.  
 RA Kintner C.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.  
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M3874; AB02039.1; -  
 DR PIR: A35844; A35844.  
 DR HSP: P00740; 1EDM.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_11.  
 DR InterPro: IPR000800; Notch.  
 DR Pfam: PF00008; ank; 6.  
 DR Pfam: PF00008; EGF; 36.  
 DR Pfam: PF00066; notch; 3.  
 DR PRINTS: PRO010; EGFBLD.  
 DR PRINTS: PRO1452; NOTCH.  
 DR SMART: SM00248; ANK; 5.  
 DR SMART: SM00179; EGF\_CA; 23.  
 DR SMART: SM00004; EGF-like; 11.  
 DR PROSITE: PS50068; ANK\_REPEAT; 4.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 23.  
 DR PROSITE: PS00022; EGF\_1; 34.  
 DR PROSITE: PS01186; EGF\_2; 29.  
 DR PROSITE: PS01187; EGF\_CA; 21.  
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Signal; Glycoprotein.  
 FT SIGNAL 1 19  
 FT CHAIN 20 2524  
 FT DOMAIN 20 1728  
 FT TRANSMEM 1729 1750  
 FT DOMAIN 1751 2524  
 FT DOMAIN 20 57  
 FT DOMAIN 58 99  
 FT DOMAIN 102 140  
 FT DOMAIN 141 177  
 FT DOMAIN 179 215  
 FT DOMAIN 217 254  
 FT DOMAIN 256 292  
 FT DOMAIN 294 332  
 FT DOMAIN 334 370  
 FT DOMAIN 371 409  
 FT DOMAIN 411 449  
 FT DOMAIN 451 487

FT DOMAIN 489 525 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 527 563 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 565 600 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 602 638 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 640 675 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 677 713 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 715 750 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 752 788 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 790 826 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 828 866 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 868 904 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 906 942 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 944 980 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 982 1018 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1020 1056 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1058 1094 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1096 1142 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1144 1180 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1182 1218 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1220 1264 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1266 1304 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1306 1346 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1347 1383 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1386 1424 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).  
FT REPEAT 1441 1478 LIN/NOTCH 1.  
FT REPEAT 1479 1520 LIN/NOTCH 2.  
FT REPEAT 1521 1560 LIN/NOTCH 3.  
FT REPEAT 1560 1599 ANK 1.  
FT REPEAT 1599 1638 ANK 2.  
FT REPEAT 1638 1677 ANK 3.  
FT REPEAT 1677 1716 ANK 4.  
FT REPEAT 1716 1755 ANK 5.  
FT REPEAT 1755 1794 ANK 6.  
FT DISULFID 22 35 BY SIMILARITY.  
FT DISULFID 35 48 BY SIMILARITY.  
FT DISULFID 48 56 BY SIMILARITY.  
FT DISULFID 56 74 BY SIMILARITY.  
FT DISULFID 74 87 BY SIMILARITY.  
FT DISULFID 87 98 BY SIMILARITY.  
FT DISULFID 98 117 BY SIMILARITY.  
FT DISULFID 117 128 BY SIMILARITY.  
FT DISULFID 128 139 BY SIMILARITY.  
FT DISULFID 139 156 BY SIMILARITY.  
FT DISULFID 156 165 BY SIMILARITY.  
FT DISULFID 165 176 BY SIMILARITY.  
FT DISULFID 176 194 BY SIMILARITY.  
FT DISULFID 194 203 BY SIMILARITY.  
FT DISULFID 203 214 BY SIMILARITY.  
FT DISULFID 214 232 BY SIMILARITY.  
FT DISULFID 232 242 BY SIMILARITY.  
FT DISULFID 242 253 BY SIMILARITY.  
FT DISULFID 253 271 BY SIMILARITY.  
FT DISULFID 271 280 BY SIMILARITY.  
FT DISULFID 280 291 BY SIMILARITY.  
FT DISULFID 291 311 BY SIMILARITY.  
FT DISULFID 311 320 BY SIMILARITY.  
FT DISULFID 320 331 BY SIMILARITY.  
FT DISULFID 331 349 BY SIMILARITY.  
FT DISULFID 349 358 BY SIMILARITY.  
FT DISULFID 358 369 BY SIMILARITY.  
FT DISULFID 369 386 BY SIMILARITY.  
FT DISULFID 386 397 BY SIMILARITY.  
FT DISULFID 397 408 BY SIMILARITY.  
FT DISULFID 408 428 BY SIMILARITY.  
FT DISULFID 428 437 BY SIMILARITY.  
FT DISULFID 437 448 BY SIMILARITY.  
FT DISULFID 448 466 BY SIMILARITY.  
FT DISULFID 466 475 BY SIMILARITY.  
FT DISULFID 475 486 BY SIMILARITY.  
FT DISULFID 486 504 BY SIMILARITY.  
FT DISULFID 504 513 BY SIMILARITY.  
FT DISULFID 513 524 BY SIMILARITY.  
FT DISULFID 524 542 BY SIMILARITY.

FT DISULFID 536 551 BY SIMILARITY.  
FT DISULFID 553 562 BY SIMILARITY.  
FT DISULFID 569 579 BY SIMILARITY.  
FT DISULFID 574 588 BY SIMILARITY.  
FT DISULFID 590 599 BY SIMILARITY.  
FT DISULFID 606 617 BY SIMILARITY.  
FT DISULFID 621 626 BY SIMILARITY.  
FT DISULFID 628 637 BY SIMILARITY.  
FT DISULFID 644 654 BY SIMILARITY.  
FT DISULFID 649 663 BY SIMILARITY.  
FT DISULFID 665 674 BY SIMILARITY.  
FT DISULFID 681 692 BY SIMILARITY.  
FT DISULFID 686 701 BY SIMILARITY.  
FT DISULFID 703 712 BY SIMILARITY.  
FT DISULFID 719 729 BY SIMILARITY.  
FT DISULFID 724 738 BY SIMILARITY.  
FT DISULFID 740 749 BY SIMILARITY.  
FT DISULFID 756 767 BY SIMILARITY.  
FT DISULFID 761 776 BY SIMILARITY.  
FT DISULFID 778 787 BY SIMILARITY.  
FT DISULFID 794 805 BY SIMILARITY.  
FT DISULFID 799 814 BY SIMILARITY.  
FT DISULFID 816 825 BY SIMILARITY.  
FT DISULFID 832 843 BY SIMILARITY.  
FT DISULFID 837 854 BY SIMILARITY.  
FT DISULFID 856 865 BY SIMILARITY.  
FT DISULFID 872 883 BY SIMILARITY.  
FT DISULFID 877 892 BY SIMILARITY.  
FT DISULFID 894 903 BY SIMILARITY.  
FT DISULFID 910 921 BY SIMILARITY.  
FT DISULFID 915 930 BY SIMILARITY.  
FT DISULFID 932 941 BY SIMILARITY.  
FT DISULFID 946 955 BY SIMILARITY.  
FT DISULFID 966 977 BY SIMILARITY.  
FT DISULFID 981 1006 BY SIMILARITY.  
FT DISULFID 1008 1017 BY SIMILARITY.  
FT DISULFID 1024 1035 BY SIMILARITY.  
FT DISULFID 1029 1044 BY SIMILARITY.  
FT DISULFID 1046 1055 BY SIMILARITY.  
FT DISULFID 1062 1073 BY SIMILARITY.  
FT DISULFID 1067 1082 BY SIMILARITY.  
FT DISULFID 1084 1093 BY SIMILARITY.  
FT DISULFID 1100 1121 BY SIMILARITY.  
FT DISULFID 1115 1130 BY SIMILARITY.  
FT DISULFID 1132 1141 BY SIMILARITY.  
FT DISULFID 1148 1159 BY SIMILARITY.  
FT DISULFID 1153 1168 BY SIMILARITY.

Query Match 11.1%; Score 223; DB 1; Length 2524;  
Best Local Similarity 23.5%; Pred. No. 5.1e-09;  
Matches 68; Conservative 20; Mismatches 101; Indels 100; Gaps 14;

Qy 126 VC-CSPGYTPDCL---ACGGGSGRPGSGNGHSGGDSRGDGSRCRHMGYGFPLCTD- 179  
Db 164 ICKPFGFHGATCKODINEC---SQNPCKNGQOCINE--FGSYRCTCQNRFTGRNCDEP 217  
Qy 180 -----CMGDFSSLNETHSTCTA-----CDSCSKRC-----SGLTN 211  
Db 218 YVPCNPSPCLNNGTCRQTDDBTSYDCTCLPGFSGQNCSEENIDDCPSNNCRNGCTCYDGVNT 277  
Qy 212 RDCGCEVGVWYLDGAGCVADVDECAAEPPCSAOPCKNANGSYTC-----EBCDSS- 262  
Db 278 YNC-QCPDWT-GQVCTEDVDVDECGQLMPNACQNGRCHMTYGGYNCVCVNGWTGEGCSENI 335  
Qy 263 --CVGCTGEGPNCKE-----CISGVAREHGQC----- 288  
Db 336 DDCANAAHSGATGCHDRVASFYCECPHGRGTGLCHLDNACISNPNCEGSSNCDTPNPNVNGKA 395  
Qy 289 -----ADVDECSIAEKTGVKKNKNCNFTPGSYVCVCPDGF 323  
Db 396 ICTCPPGYTGACNNDVDECSLGANPC--EHGRCRTNVLGFSQCNCPQGT 443  
RESULT 27

ID	EMBL_HUMAN	STANDARD:	PRT:	886 AA.
AC	014245:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Cell surface glycoprotein EMR1 precursor (EMR1 hormone receptor).			
GN	EMR1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Carnivora; Hominiinae; Homo.			
OX	NCBI_TaxID=9606;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95324926; PubMed=7601460;			
RA	Baud V., Chissoc S.L., Viegas-Pequignot E., Diliong S., Nguyen V.C.,			
RA	Roe B.A., Lipinski M.,			
RT	"EMR1, an unusual member in the family of hormone receptors with			
RL	seven transmembrane segments.";			
CC	Genomics 26:334-344(1995).			
CC	-1- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR			
CC	AN INTERACTION WITH A PROTEIN LIGAND.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN			
CC	PERIPHERAL BLOOD MONONUCLEAR CELLS.			
CC	-1- PTM: N- AND O-GLYCOSYLATED; (POSSIBLE).			
CC	-1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-1- SIMILARITY: CONTAINS 1 GPS DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X81479; CA57232.1; .			
DR	HSSP; P35555; IEMN.			
DR	MIM; 600493; .			
DR	InterPro: IPR000152; Asx_hydroxyl.			
DR	InterPro: IPR000561; EGF-like.			
DR	InterPro: IPR001881; EGF_Ca.			
DR	InterPro: IPR000832; GPCR_secretin.			
DR	InterPro: IPR000203; PKD_cys_rich.			
DR	Pfam; PF00002; 7tm_2; 1.			
DR	Pfam; PF00008; EGF; 5.			
DR	Pfam; PF01825; GPS; 1.			
DR	PRINTS; PR01128; EMRHOMONER.			
DR	SMART; SMART0179; EGF_CA_5.			
DR	SMART; SMART0001; EGF_like; 1.			
DR	SMART; SMART0303; GPS; 1.			
DR	PROSITE; PS00010; ASX_HYDROXYL; 6.			
DR	PROSITE; PS01186; EGF_2; 2.			
DR	PROSITE; PS01187; EGF_CA; 5.			
DR	PROSITE; PS02221; GPS; 1.			
DR	PROSITE; PS00650; G_PROTEIN_REC_P2_2; 1.			
DR	PROSITE; PS02621; G_PROTEIN_REC_P2_4; 1.			
KW	G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;			
FT	EGF-like domain; Repeat; Signal.			
FT	SIGNAL	1	17	POTENTIAL.
FT	CHAIN	18	886	CELL SURFACE GLYCOPROTEIN EMR1.
FT	DOMAIN	18	599	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	600	627	POTENTIAL.
FT	DOMAIN	628	634	CYTOSOLASMIC (POTENTIAL).
FT	TRANSMEM	635	656	POTENTIAL.
FT	DOMAIN	657	666	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	667	690	POTENTIAL.
FT	DOMAIN	691	709	CYTOSOLASMIC (POTENTIAL).
FT	TRANSMEM	710	731	POTENTIAL.
FT	DOMAIN	732	747	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	748	776	POTENTIAL.

Query Match	Best Local Similarity	Score	DB 1	Length	886:
Matches	79:	Conservative	20:	Mismatches	95:
				Indels	89:
					Gaps
130	PETYPDCLACGGGQPCSGNGHCSGGDSROGGDSRCRMGYOGLCTDCM---	DGYTS	186		
Db	74	PCVPRCKDIDEC--SOSPPOPCGPNSSCK--NLSGRKYSCLDFFSSPTGNDWVPGKGNFS	129		
147	SLRNETHSICTCADSDSKCKSGLTNRDGE-----	CEVGVNLDGACVDV	232		
Db	130	-----CTDINE---C--LTISVCPESHSDCVNMSGSTSCQVGFISRNSTCEDVN	174		
233	ECAAEPSPCSAOPCKNANGSYTC-----	EECDSSC--V	264		
Db	175	EC-ADPRACPEHATCNNTNMGVNSYSCPNCFESSSGHLSCQGLKASCEDIDECTEMCPINS	233		
265	GCTGGSPGN--CKEISGTYAREHQ-----	CADVDEGLAETCYRKNECYNTPG	313		
Db	234	TCTNT-PSISYFC-TCHPGAPSSQILNFDGVECRDIDECQDDPSTC-GPNSICTNALG	290		
314	SVYCVCPDGFETEDACVPAEAFEGESPPQ---	LSREPL	353		
Db	291	SYSCGCIYGFH-----PNEGSGQKDGNFSCQRLVFKCKREY	326		
TRANSMEM	777	794			
DOMAIN	795	814			
DOMAIN	815	829			
TRANSMEM	830	852			
DOMAIN	853	886			
DOMAIN	31	79			
DOMAIN	80	131			
DOMAIN	132	171			
DOMAIN	172	213			
DOMAIN	221	267			
DOMAIN	268	316			
DOMAIN	547	596			
DOMAIN	317	599			
DISULFID	35	47			
DISULFID	41	56			
DISULFID	58	78			
DISULFID	84	97			
DISULFID	91	106			
DISULFID	108	130			
DISULFID	136	148			
DISULFID	142	157			
DISULFID	159	170			
DISULFID	176	188			
DISULFID	182	197			
DISULFID	199	212			
DISULFID	225	235			
DISULFID	229	244			
DISULFID	246	266			
DISULFID	272	285			
DISULFID	279	294			
DISULFID	296	315			
CARBOHYD	94	94			
CARBOHYD	99	99			
CARBOHYD	127	127			
CARBOHYD	167	167			
CARBOHYD	189	189			
CARBOHYD	194	194			
CARBOHYD	232	232			
CARBOHYD	258	258			
CARBOHYD	312	312			
CARBOHYD	366	366			
CARBOHYD	375	375			
CARBOHYD	448	448			
CARBOHYD	661	661			
SEQUENCE	886 AA:	97680 MW:	7456CA56FB624D99	CRC64:	



AC	000918-1993 (Rel. 26, Created)
DT	01-JUL-1993 (Rel. 26, Last sequence update)
DT	01-OCT-1996 (Rel. 34, Last annotation update)
DE	Latent transforming growth factor beta binding protein 1 precursor
DE	(Transforming growth factor beta-1 binding protein 1) (TGF-beta1-BP-
GN	1) (Transforming growth factor beta-1 masking protein, large subunit).
GN	Ltbp1.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Muriinae; Rattus.
OX	NCBI_Taxid=10116;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX	MEDLINE=91062373; PubMed=2247454;
RA	Tsuji T., Okada F., Yamaguchi K., Nakamura T.;
RT	"Molecular cloning of the large subunit of transforming growth factor
RT	type beta masking protein and expression of the mRNA in various rat
RL	tissues".
RL	Proc. Natl. Acad. Sci. U.S.A. 87:8835-8839(1990).
CC	-I- SUBUNIT: THE LARGE LATENT COMPLEX OF TGF-BETA1 FROM PLATELETS IS
CC	COMPOSED OF THE TGF-BETA1 MOLECULE NONCOVALENTLY ASSOCIATED WITH
CC	A MASKING PROTEIN CONSISTING OF A DISULFIDE-BONDED COMPLEX OF A
CC	DIMER OF THE N-TERMINAL PROPEPTIDE OF THE TGF-BETA1 PRECURSOR AND
CC	A THIRD COMPONENT DENOTED TGF-BETA1-BP (OR MP LARGE SUBUNIT).
CC	TGF-BETA1-BP DOES NOT BIND DIRECTLY TO ACTIVE TGF-BETA1.
CC	-I- SIMILARITY: CONTAINS 18 EGF-LIKE DOMAINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; M55431; AAAA42235.1; -.
DR	PIR; A38261; A38261.
DR	HSSP; P35355; IEMN.
DR	InterPro; IPRO000152; Asx_hydroxy1.
DR	InterPro; IPRO000561; EGF_1like.
DR	InterPro; IPRO01881; EGF_Ca.
DR	InterPro; IPRO02212; TB.
DR	Pfam; PF00008; EGF; 16.
DR	Pfam; PF00683; TB; 4.
DR	SMART; SMO0179; EGF_CA; 13.
DR	SMART; SMO0001; EGF_1like; 5.
DR	PROSITE; PS00010; ASX_HYDROXYL; 13.
DR	PROSITE; PS00022; EGF_1; 2.
DR	PROSITE; PS01186; EGF_2; 10.
DR	PROSITE; PS01187; EGF_CA; 15.
KW	Growth factor binding; Repeat; EGF-like domain; Hydroxylation; Signal.
FT	SIGNAL
FT	PROPEP
FT	CHAIN
FT	1       21       736
FT	737      1577
FT	SITE          734      736
FT	SITE          1575    1577
FT	PROPEP        1578    1712
FT	DOMAIN        181      213
FT	DOMAIN        391      423
FT	REPEAT        551      604
FT	DOMAIN        618      658
FT	REPEAT        721      721
FT	DOMAIN        865      906
FT	DOMAIN        907      948
FT	DOMAIN        949      989
FT	DOMAIN        990      1029
FT	DOMAIN        1030    1070
FT	DOMAIN        1071    1111
FT	DOMAIN        1112    1152
FT	DOMAIN        1153    1193
FT	DOMAIN        1194    1235
FT	DOMAIN        1235
FT	LATENT TRANSFORMING GROWTH FACTOR BETA
FT	BINDING PROTEIN 1.
FT	CLEAVAGE (POTENTIAL).
FT	CLEFTAGE (POTENTIAL).
FT	POTENTIAL.
FT	EGF-LIKE 1.
FT	INTERNAL REPEAT 1.
FT	EGF-LIKE 2.
FT	INTERNAL REPEAT 1.
FT	EGF-LIKE 3.
FT	CALCIUM-BINDING (POTENTIAL).
FT	INTERNAL REPEAT 2.
FT	EGF-LIKE 4.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 5.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 6.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 7.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 8.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 9.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 10.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 11.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 12.
FT	CALCIUM-BINDING (POTENTIAL).
FT	EGF-LIKE 12.
FT	CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	1236	1277	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1278	1320	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1340	1392	INTERNAL REPEAT 3
FT	DOMAIN	1415	1457	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1458	1498	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1517	1568	INTERNAL REPEAT 4
FT	DOMAIN	1612	1652	EGF-LIKE 17
FT	DOMAIN	1653	1697	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	185	195	BY SIMILARITY.
FT	DISULFID	189	201	BY SIMILARITY.
FT	DISULFID	203	212	BY SIMILARITY.
FT	DISULFID	395	405	BY SIMILARITY.
FT	DISULFID	399	411	BY SIMILARITY.
FT	DISULFID	413	422	BY SIMILARITY.
FT	DISULFID	622	633	BY SIMILARITY.
FT	DISULFID	628	642	BY SIMILARITY.
FT	DISULFID	644	657	BY SIMILARITY.
FT	DISULFID	869	881	BY SIMILARITY.
FT	DISULFID	876	890	BY SIMILARITY.
FT	DISULFID	892	905	BY SIMILARITY.
FT	DISULFID	911	923	BY SIMILARITY.
FT	DISULFID	918	932	BY SIMILARITY.
FT	DISULFID	934	947	BY SIMILARITY.
FT	DISULFID	953	964	BY SIMILARITY.
FT	DISULFID	959	973	BY SIMILARITY.
FT	DISULFID	976	988	BY SIMILARITY.
FT	DISULFID	994	1005	BY SIMILARITY.
FT	DISULFID	1000	1014	BY SIMILARITY.
FT	DISULFID	1017	1028	BY SIMILARITY.
FT	DISULFID	1034	1045	BY SIMILARITY.
FT	DISULFID	1040	1054	BY SIMILARITY.
FT	DISULFID	1056	1069	BY SIMILARITY.
FT	DISULFID	1075	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1110	BY SIMILARITY.
FT	DISULFID	1116	1127	BY SIMILARITY.
FT	DISULFID	1122	1136	BY SIMILARITY.
FT	DISULFID	1138	1151	BY SIMILARITY.
FT	DISULFID	1157	1169	BY SIMILARITY.
FT	DISULFID	1164	1178	BY SIMILARITY.
FT	DISULFID	1180	1192	BY SIMILARITY.
FT	DISULFID	1198	1210	BY SIMILARITY.
FT	DISULFID	1204	1219	BY SIMILARITY.
FT	DISULFID	1221	1234	BY SIMILARITY.
FT	DISULFID	1240	1252	BY SIMILARITY.
FT	DISULFID	1246	1261	BY SIMILARITY.
FT	DISULFID	1263	1276	BY SIMILARITY.
FT	DISULFID	1282	1294	BY SIMILARITY.
FT	DISULFID	1289	1303	BY SIMILARITY.
FT	DISULFID	1305	1319	BY SIMILARITY.
FT	DISULFID	1419	1432	BY SIMILARITY.
FT	DISULFID	1427	1441	BY SIMILARITY.
FT	DISULFID	1443	1456	BY SIMILARITY.
FT	DISULFID	1462	1473	BY SIMILARITY.
FT	DISULFID	1468	1482	BY SIMILARITY.
FT	DISULFID	1484	1497	BY SIMILARITY.
FT	DISULFID	1616	1627	BY SIMILARITY.
FT	DISULFID	1622	1636	BY SIMILARITY.
FT	DISULFID	1638	1651	BY SIMILARITY.
FT	DISULFID	1657	1672	BY SIMILARITY.
FT	DISULFID	1687	1681	BY SIMILARITY.
FT	DISULFID	1683	1696	BY SIMILARITY.
FT	CARBOHYD	339	339	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	370	370	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	416	416	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	612	612	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1042	1042	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1242	1242	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1357	1357	N-LINKED (GLCNAC. . .) (POTENTIAL).
90	SEQUENCE	1712	186598	MM;1650BDEAA691FDJ34 CRC64;

Query match

11.187

Score 222; DB 1; Length 1712;



```

FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 867 BY SIMILARITY.
FT DISULFID 874 885 BY SIMILARITY.
FT DISULFID 879 894 BY SIMILARITY.
FT DISULFID 912 923 BY SIMILARITY.
FT DISULFID 917 932 BY SIMILARITY.
FT DISULFID 934 943 BY SIMILARITY.
FT DISULFID 988 999 BY SIMILARITY.
FT DISULFID 993 1008 BY SIMILARITY.
FT DISULFID 1010 1019 BY SIMILARITY.
FT DISULFID 1026 1037 BY SIMILARITY.

```

```

FT DISULFID 1031 1046 BY SIMILARITY.
FT DISULFID 1048 1057 BY SIMILARITY.
FT DISULFID 1064 1075 BY SIMILARITY.

Query Match 11.1%; Score 222; DB 1; Length 2444;
Best Local Similarity 24.6%; Pred. No. 5.9e-09;
Matches 75; Conservative 26; Mismatches 104; Indels 100; Gaps 15;

QY 125 KVCSPGTY-GPDCACGGSGRPGSGNGHCS-GDSSRGGDSCRMGOGPLCTDCMD 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 44 EACVCGGATVPGRCQDPNPLCLSPCKMAGTCHYVDRGVADYACSCALFSGPLCLTPD 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 GYFSSLRNE-----THSICTACDESCKTCSTGLTRNDGCE---CEGVWLDGACV----- 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 N-ACLTNCRNCGTDDLLTFEYKCRCPGWSGKSCQADPCASNPANGGCLPFEAS 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 -----DVECAAEPPPCSAOPCKANANSYTC----- 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 YICHCPPSFHPTCRQDVNDCGKPRLCRHGGTCNHEVGSYRCVCRATHTGNCERPYP 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 -----EECDSSCV-GCTGEG-----PG-NCKE---CISGYANE 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 CSRSPQNGTCKPTGDTVTHEC--ACLPFTGNCSEINDDCGRNCKRGACVDSVNTY 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 HGQC-----ADVDCSLAEKTCVRKNENCVNTPGSIYVCPDGF-----EETEDA 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 NCPCPPEMTGXCTEDVDQCQMPNAC--ONGGTCNTHGYNVCVNGWTGECDSINID 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 330 CVPPA 334
Db 339 CASAA 343

RESULT 31
NIDO_HUMAN STANDARD; PRT; 1247 AA.
ID NIDO_HUMAN
AC P14543; O14942;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nidogen precursor (Entactin).
GN NID.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90091745; PubMed=2574658;
RA Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,
RA Chu M.-L., Knowlton R., Mann K., Deutzmann R., Timpl R., Uitto J.;
RT "Human nidogen: complete amino acid sequence and structural domains
RT deduced from cDNAs, and evidence for polymorphism of the gene.";
RL DNA 8:581-594(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96044428; PubMed=7557988;
RA Zimmermann K., Holschen S., Hafner M., Nischt R.;
RT "Genomic sequences and structural organization of the human nidogen
RT gene (NID).";
RL Genomics 27:245-250(1995).
RN [3]
RP SEQUENCE OF 667-1247 FROM N.A.
RX TISSUE=Placenta;
RA MEDLINE=89270475; PubMed=2471408;
RA Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Passage E., Weil D.,
RA Timpl R., Chu M.-L., Uitto J.;
RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the
RT gene to chromosome 14q3.1";
RL Am. J. Hum. Genet. 44:876-885(1989).
CC -!- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN
CC BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.
CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-

```



CC EXTRACELLULAR MATRIX INTERACTIONS.  
 CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.  
 CC -1- PTM: N- AND O-GLYCOSYLATED.  
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M30269; AA55993.1; -  
 DR EMBL: X82245; CA557709.1; JOINED.  
 DR EMBL: X84819; CA557709.1; JOINED.  
 DR EMBL: X84820; CA557709.1; JOINED.  
 DR EMBL: X84821; CA557709.1; JOINED.  
 DR EMBL: X84822; CA557709.1; JOINED.  
 DR EMBL: X84823; CA557709.1; JOINED.  
 DR EMBL: X84824; CA557709.1; JOINED.  
 DR EMBL: X84825; CA557709.1; JOINED.  
 DR EMBL: X84826; CA557709.1; JOINED.  
 DR EMBL: X84827; CA557709.1; JOINED.  
 DR EMBL: X84828; CA557709.1; JOINED.  
 DR EMBL: X84829; CA557709.1; JOINED.  
 DR EMBL: X84830; CA557709.1; JOINED.  
 DR EMBL: X84831; CA557709.1; JOINED.  
 DR EMBL: X84832; CA557709.1; JOINED.  
 DR EMBL: X84833; CA557709.1; JOINED.  
 DR EMBL: X84834; CA557709.1; JOINED.  
 DR EMBL: X84835; CA557709.1; JOINED.  
 DR EMBL: X84836; CA557709.1; JOINED.  
 DR EMBL: X84837; CA557709.1; JOINED.  
 DR EMBL: M27445; AA557261.1; -  
 DR PIR: A33322; MMHUND.  
 DR HSSP: P07204; IADX.  
 DR MIM: 131390; -  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_CA.  
 DR InterPro: IPR000033; Ldl\_receptor\_rep.  
 DR InterPro: IPR003886; Nidogen\_ext.  
 DR InterPro: IPR000716; Thyroglobulin\_1.  
 DR Pfam: PF00008; EGF\_6.  
 DR Pfam: PF00058; Ldl\_receptor\_b; 3.  
 DR Pfam: PF00086; thyroglobulin\_1; 1.  
 DR SMART: SM00179; EGF\_CA; 2.  
 DR SMART: SM00001; EGF\_Like; 4.  
 DR SMART: SM00135; LY; 5.  
 DR SMART: SM00539; NIDO; 1.  
 DR SMART: SM00211; TY; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1; 1.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; 5.  
 DR PROSITE: PS01187; EGF\_CA; 2.  
 KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation;  
 KW Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1247 NIDOGEN,  
 FT DOMAIN 29 669 I (LARGER GLOBULAR DOMAIN).  
 FT DOMAIN 670 917 II (CYSTEINE-RICH).  
 FT DOMAIN 918 1247 III (SMALLER GLOBULAR DOMAIN).  
 FT DOMAIN 386 426 EGF-LIKE 1.  
 FT DOMAIN 668 709 EGF-LIKE 2.  
 FT DOMAIN 710 751 EGF-LIKE 3.  
 FT DOMAIN 758 801 EGF-LIKE 4.  
 FT DOMAIN 802 840 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 872 919 THYROGLOBULIN TYPE I.  
 FT DOMAIN 989 1030 LDL-RECEPTOR YWTD MOTIF 1.

FT DOMAIN 1032 1073 LDL-RECEPTOR YWTD MOTIF 2.  
 FT DOMAIN 1075 1118 LDL-RECEPTOR YWTD MOTIF 3.  
 FT DOMAIN 1124 1163 LDL-RECEPTOR YWTD MOTIF 4.  
 FT DOMAIN 1208 1244 EGF-LIKE 6.  
 FT MOD\_RES 289 289 SULFATION (POTENTIAL).  
 FT MOD\_RES 296 296 SULFATION (POTENTIAL).  
 FT DISULFID 672 685 BY SIMILARITY.  
 FT DISULFID 679 695 BY SIMILARITY.  
 FT DISULFID 697 708 BY SIMILARITY.  
 FT DISULFID 714 727 BY SIMILARITY.  
 FT DISULFID 721 736 BY SIMILARITY.  
 FT DISULFID 738 750 BY SIMILARITY.  
 FT DISULFID 762 777 BY SIMILARITY.  
 FT DISULFID 1219 1232 BY SIMILARITY.  
 FT DISULFID 1234 1243 BY SIMILARITY.  
 FT SITE 702 704 CELL ATTACHMENT SITE.  
 FT CARBOHYD 1137 1137 N-LINKED (GLCNAC: . .) (POTENTIAL).  
 FT CONFLICT 33 34 EL -> SS (IN REF. 2).  
 FT CONFLICT 37 42 FGPGQG -> SARDR (IN REF. 2).  
 FT CONFLICT 1115 1115 H -> T (IN REF. 3).  
 SQ SEQUENCE 1247 AA; 136488 MW; 4681B5B3CEC15758 CAC64;  
 Query Match 11.0%; Score 221; DB 1; Length 1247;  
 Best Local Similarity 27.1%; Pred. No. 3,7e-09;  
 Matches 54; Conservative 28; Mismatches 75; Indels 42; Gaps 7;  
 QY 174 GPLCTDMGTFSSLRNETHSICACDESCKTCGLTNRDGEGEYGVWLDGACVYDE 233  
 Db 658 GPV-----REGSPDALQNPICYIGTGCCTNMACRGPPTQFCESIGFRGDRCYIDE 713  
 QY 234 CAAPPPCSAQAQFCNNNGSYTCEC-----DSSCVGTGGRP-----GNCK-- 275  
 Db 714 CSEQPSVCGSHTCNNHPGFRFC-ECVEGYOFSEGCVAVVDPRPNTYCTGHLNDIP 772  
 QY 276 -----ECISGYAREHGOCADVDECSLAEKTCVRNENCYNTPGSYVCVP 320  
 Db 773 QRAQCIVTGGSSYSCSLPFGSGDQAQDQVDEC---QPSRCHPDACVNTPGSFTGCK 829  
 QY 321 DGFETEDACYPAPAEAT 339  
 Db 830 PGYGDGFRCV-PGEVERKT 847  
 RESULT 32  
 NCCL\_RAT  
 ID NCCL\_RAT STANDARD; PRT; 2531 AA.  
 AC 007008;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neurogenic locus notch homolog protein 1 precursor.  
 GN NCCHL.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Schwann cell;  
 RX MEDLINE=92111383; PubMed=1764995;  
 RA Weinmaster G., Roberts V.J., Lemke G.;  
 RT "A homolog of Drosophila Notch expressed during mammalian  
 development";  
 RL Development 113:199-205(1991).  
 [2]  
 REVIEWS TO 1652-1653.



Query Match 11.08; Score 220; DB 1; Length 2531;  
 Best Local Similarity 23.6%; Pred. No. 8; 5e-09;  
 Matches 76; Conservative 23; Mismatches 97; Indels 126; Gaps 17;

Query	Similarity	Score	DB	Length	2531
126 VC-CSPGTYPPDCL-ACGGGSGRP--CSGNHSGSDGSRDGGSCRCRMGYGFLC-177	11.08	220	DB 1	2531	
163 ICGCPHFHPTCRQVNYEC--SQNGLCRHGCTCNE--ISYRCACRAHTTGPHE-216	11.08	220	DB 1	2531	
178 -----TDCMDGYFSSLRNETHSICTA-----CDESKTCGSLTNRDCECEVGW-221	11.08	220	DB 1	2531	
217 LPVYPCSPFCQNGGTCRPTGDTTNEACALRGFAGQNCCEWVDPCG-----NKNANG-269	11.08	220	DB 1	2531	
222 VLDEGACV-----DVDECAEPFPCSAOFCCKNANGSYTC-----256	11.08	220	DB 1	2531	
270 -----GACVDGVNVTNCRCPPEMTGQYCTEDVDCCOLMPNACQNNAGTCHNSHGYNVCVN-325	11.08	220	DB 1	2531	
257 -----EECDSSCGVCGTGBG--PGNCKE-----CISGYAREHG-286	11.08	220	DB 1	2531	
326 GMTGEDSCDNIIDPCASACFOGATCHDRVASFYCECPHGRTGLCHLNDACISNPNCEGS-385	11.08	220	DB 1	2531	
287 QC-----ADVDECSLAETKCVKRNENGYNTPGSYVCVCPDGF-323	11.08	220	DB 1	2531	
386 NCDPNPVNGKALICTCPRGYTGAPACSDVDVDCALCANPCENHAK-CLMTLGSFECQCLGY-444	11.08	220	DB 1	2531	
324 -----EETEDACVP-PAEAEAT-339	11.08	220	DB 1	2531	
445 TGPRCEIDVNECISNCONDAT-466	11.08	220	DB 1	2531	

RESULT 33  
 LML2\_CABEL  
 ID LML2\_CABEL STANDARD; PRT; 3672 AA.  
 AC Q21313;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE laminin-like protein K08C7.3 precursor.  
 GN K08C7.3.  
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiidae; Rhabditoidea;  
 OC Rhabditiidae; Pelodetidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Berks M.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1ab-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; Z70286; CA94293.1; --  
 DR HSSP; P02468; 1KLO.  
 DR WormPep; K08C7.3; CE06136.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001886; LamNT.  
 DR InterPro; IPR000034; laminin\_B.  
 DR InterPro; IPR002049; laminin\_EGF.  
 DR InterPro; IPR001791; laminin\_G.  
 DR Pfam; PF00052; laminin\_B; 1.  
 DR Pfam; PF00053; laminin\_EGF; 21.  
 DR Pfam; PF00054; laminin\_G; 5.  
 DR Pfam; PF00055; laminin\_Nterm; 1.  
 DR PRINTS; PR00011; EGF/LAMININ.  
 DR PRODOM; PD002082; LamNT; 1.  
 DR PRODOM; PD003031; laminin\_B; 1.  
 DR SMART; SM00180; EGF; 21.  
 DR SMART; SM00281; Lamb; 1.  
 DR SMART; SM00282; Lamb; 5.  
 DR SMART; SM00136; LamNT; 1.  
 DR PROSITE; PS00022; EGF\_1; 19.  
 DR PROSITE; PS0186; EGF\_2; 4.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 21.  
 DR PROSITE; PS50025; LAM G DOMAIN; 5.  
 DR KW Hypothetical protein; laminin EGF-like domain; Signal; Repeat.  
 FT CHAIN 1..27  
 FT SIGNAL 1..27  
 FT CHAIN 28..3672  
 FT DOMAIN 28..297 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 298..356 LAMININ EGF-LIKE 1.  
 FT DOMAIN 357..426 LAMININ EGF-LIKE 2.  
 FT DOMAIN 427..471 LAMININ EGF-LIKE 3.  
 FT DOMAIN 472..518 LAMININ EGF-LIKE 4.  
 FT DOMAIN 519..563 LAMININ EGF-LIKE 5.  
 FT DOMAIN 564..609 LAMININ EGF-LIKE 6.  
 FT DOMAIN 610..655 LAMININ EGF-LIKE 7.  
 FT DOMAIN 656..700 LAMININ EGF-LIKE 8.  
 FT DOMAIN 701..755 LAMININ EGF-LIKE 9.  
 FT DOMAIN 756..808 LAMININ EGF-LIKE 10.  
 FT DOMAIN 809..839 LAMININ EGF-LIKE 11 (INCOMPLETE).  
 FT DOMAIN 1415..1460 LAMININ EGF-LIKE 12.  
 FT DOMAIN 1461..1505 LAMININ EGF-LIKE 13.  
 FT DOMAIN 1506..1553 LAMININ EGF-LIKE 14.  
 FT DOMAIN 1554..1604 LAMININ EGF-LIKE 15.  
 FT DOMAIN 1605..1614 LAMININ EGF-LIKE 16 (N-TERMINAL).  
 FT DOMAIN 1615..1796 LAMININ DOMAIN IV.  
 FT DOMAIN 1797..1829 LAMININ EGF-LIKE 16 (C-TERMINAL).  
 FT DOMAIN 1830..1879 LAMININ EGF-LIKE 17.  
 FT DOMAIN 1880..1936 LAMININ EGF-LIKE 18.  
 FT DOMAIN 1937..1989 LAMININ EGF-LIKE 19.  
 FT DOMAIN 1990..2036 LAMININ EGF-LIKE 20.  
 FT DOMAIN 2037..2083 LAMININ EGF-LIKE 21.  
 FT DOMAIN 2084..2131 LAMININ EGF-LIKE 22.  
 FT DOMAIN 2693..2884 LAMININ G-LIKE 1.

FT	DOMAIN	2896	3066	LAMININ G-LIKE 2.
FT	DOMAIN	3072	3235	LAMININ G-LIKE 3.
FT	DOMAIN	3310	3482	LAMININ G-LIKE 4.
FT	DOMAIN	3488	3669	LAMININ G-LIKE 5.
FT	DISULFID	298	307	BY SIMILARITY.
FT	DISULFID	300	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	334	354	BY SIMILARITY.
FT	DISULFID	357	366	BY SIMILARITY.
FT	DISULFID	359	391	BY SIMILARITY.
FT	DISULFID	394	403	BY SIMILARITY.
FT	DISULFID	406	424	BY SIMILARITY.
FT	DISULFID	427	438	BY SIMILARITY.
FT	DISULFID	429	445	BY SIMILARITY.
FT	DISULFID	447	456	BY SIMILARITY.
FT	DISULFID	459	469	BY SIMILARITY.
FT	DISULFID	472	484	BY SIMILARITY.
FT	DISULFID	474	491	BY SIMILARITY.
FT	DISULFID	493	502	BY SIMILARITY.
FT	DISULFID	505	516	BY SIMILARITY.
FT	DISULFID	519	531	BY SIMILARITY.
FT	DISULFID	521	538	BY SIMILARITY.
FT	DISULFID	540	549	BY SIMILARITY.
FT	DISULFID	552	561	BY SIMILARITY.
FT	DISULFID	564	576	BY SIMILARITY.
FT	DISULFID	566	583	BY SIMILARITY.
FT	DISULFID	585	594	BY SIMILARITY.
FT	DISULFID	597	607	BY SIMILARITY.
FT	DISULFID	610	622	BY SIMILARITY.
FT	DISULFID	612	629	BY SIMILARITY.
FT	DISULFID	631	640	BY SIMILARITY.
FT	DISULFID	643	653	BY SIMILARITY.
FT	DISULFID	656	668	BY SIMILARITY.
FT	DISULFID	658	676	BY SIMILARITY.
FT	DISULFID	676	685	BY SIMILARITY.
FT	DISULFID	688	698	BY SIMILARITY.
FT	DISULFID	701	715	BY SIMILARITY.
FT	DISULFID	703	724	BY SIMILARITY.
FT	DISULFID	726	735	BY SIMILARITY.
FT	DISULFID	738	753	BY SIMILARITY.
FT	DISULFID	756	770	BY SIMILARITY.
FT	DISULFID	758	777	BY SIMILARITY.
FT	DISULFID	779	788	BY SIMILARITY.
FT	DISULFID	791	806	BY SIMILARITY.
FT	DISULFID	1415	1427	BY SIMILARITY.
FT	DISULFID	1417	1434	BY SIMILARITY.
FT	DISULFID	1436	1445	BY SIMILARITY.
FT	DISULFID	1448	1458	BY SIMILARITY.
FT	DISULFID	1461	1469	BY SIMILARITY.
FT	DISULFID	1463	1476	BY SIMILARITY.
FT	DISULFID	1478	1487	BY SIMILARITY.
FT	DISULFID	1490	1503	BY SIMILARITY.
FT	DISULFID	1506	1520	BY SIMILARITY.
FT	DISULFID	1508	1527	BY SIMILARITY.
FT	DISULFID	1529	1538	BY SIMILARITY.
FT	DISULFID	1541	1551	BY SIMILARITY.
FT	DISULFID	1554	1566	BY SIMILARITY.
FT	DISULFID	1575	1584	BY SIMILARITY.
FT	DISULFID	1587	1602	BY SIMILARITY.
FT	DISULFID	1830	1839	BY SIMILARITY.
FT	DISULFID	1832	1846	BY SIMILARITY.
FT	DISULFID	1849	1858	BY SIMILARITY.
FT	DISULFID	1861	1877	BY SIMILARITY.
FT	DISULFID	1880	1894	BY SIMILARITY.
FT	DISULFID	1882	1905	BY SIMILARITY.
FT	DISULFID	1907	1916	BY SIMILARITY.
FT	DISULFID	1919	1934	BY SIMILARITY.
FT	DISULFID	1937	1951	BY SIMILARITY.
FT	DISULFID	1939	1958	BY SIMILARITY.
FT	DISULFID	1961	1970	BY SIMILARITY.
FT	DISULFID	1973	1987	BY SIMILARITY.
FT	DISULFID	1990	2000	BY SIMILARITY.

FT	DISULFID	1992	2007	BY SIMILARITY.
FT	DISULFID	2009	2018	BY SIMILARITY.
FT	DISULFID	2021	2034	BY SIMILARITY.
FT	DISULFID	2037	2048	BY SIMILARITY.
FT	DISULFID	2039	2055	BY SIMILARITY.
FT	DISULFID	2057	2066	BY SIMILARITY.
FT	DISULFID	2069	2081	BY SIMILARITY.
FT	DISULFID	2084	2096	BY SIMILARITY.
FT	DISULFID	2103	2103	BY SIMILARITY.
FT	DISULFID	2105	2114	BY SIMILARITY.
FT	DISULFID	2117	2129	BY SIMILARITY.
FT	CARBOHYD	121	121	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	140	140	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	249	249	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	351	351	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	477	477	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	511	511	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	634	634	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	761	761	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1014	1014	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1705	1705	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1756	1756	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1868	1868	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1944	1944	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1966	1966	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2002	2002	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2159	2159	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2207	2207	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2231	2231	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2235	2235	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2401	2401	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2421	2421	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2487	2487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2821	2821	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3087	3087	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3242	3242	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3541	3541	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	3672 AA;	404223 MW;	28E262DB5F14BFA CRC64;

Query Match 10.8%; Score 217.5; DB 1; Length 3672;  
Best Local Similarity 23.1%; Pred. No. 1.9e-08;  
Matches 93; Conservative 31; Mismatches 170; Indels 109; Gaps 20;

QY	31	CHR-CRGLVDKRNQGVADAKKN-----FGGGTAW--EKKTLK-----YESS	71
DB	331	CERCPCGFYQK--OMQAAIAHNNFTCEACNCFGRSNECEYDAEVLNKSIDSQGYEGG	388
QY	72	EIRLLEILGLCE----SSDECNOMLEAOEHLBAMW----LQKSEYPLDFEMFCV	121
DB	389	-----GYCKNCRENTEGVNCKNSCFGYFPRBEGVWNNEDPQCKVCCDDPKHKGACA	439
QY	122	KTLKVC-CSPGTYGPPLCAGG-----SORPCSGNGHS-GDGSNGSGSCRHMGYOG	174
DB	440	EETGKCECPFRFVGEDDCDCASGYDAPCKPCECVNNGTIGDVLPEBGQCPCKAGFGG	499
QY	175	PLCTDMDYFESSLRNETHSICACDESKCTSGGLNR-----DCGEVEGVWLD	224
DB	500	TYCETCADGTWNTACGVCVCDATGSEHNGCSASTGQCECKPATYAGLSCDKQVGYFED	559
QY	225	EGACVVDVDECAEPPPSAAQ--FCKNANGSYTEEC-----SSCVGTGEGBG--	272
DB	560	DCFKNCDDPMGTGGVCDQDTTGGCLCKBGPAGBKORROIARYGYNCAACADGAGTIS	619
QY	273	-----NKEICISGY-----AREHQCADVD-EC-----S	295
DB	620	PECDATSGOCPCNGNFTGRTCDKCAAGFYNDPCRCCECLLGAKGOTCDSDNGQCYCKGN	679
QY	296	LAETGVARNENGYNPGSYVCV--PDGFETEDAC--VPAE	335
DB	680	FGEGRCDRCCKPNFYNFPICEBCNCPNSGVTRDFQGDCKVSPGE	722



```

DR PIR: A05267; A05267.
DR HSPB; P00740; 11XA.
DR FLYBASE; Fbn0004647; N.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF000023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00101; EGFBLDOD.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 4.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00001; EGF_Like; 13.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 28.
DR PROSITE; PS01187; EGF_CA; 22.
DR Transmembrane; Signal; Glycoprotein.
KW DIFFERENTIATION; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
FT SIGNAL 1 44
FT CHAIN 45 2703
FT DOMAIN 45 1745
FT TRANSMEM 1746 1766
FT DOMAIN 1767 2703
FT DOMAIN 58 95
FT DOMAIN 96 136
FT DOMAIN 139 176
FT DOMAIN 177 215
FT DOMAIN 217 253
FT DOMAIN 255 291
FT DOMAIN 293 329
FT DOMAIN 331 370
FT DOMAIN 372 408
FT DOMAIN 409 447
FT DOMAIN 449 486
FT DOMAIN 488 524
FT DOMAIN 526 562
FT DOMAIN 564 600
FT DOMAIN 602 637
FT DOMAIN 639 675
FT DOMAIN 677 713
FT DOMAIN 715 751
FT DOMAIN 753 789
FT DOMAIN 791 827
FT DOMAIN 829 865
FT DOMAIN 867 905
FT DOMAIN 907 944
FT DOMAIN 946 982
FT DOMAIN 984 1020
FT DOMAIN 1022 1058
Query Match 10.8%; Score 216.5; DB 1; Length 2703;
Best Local Similarity 23.3%; Pred. No. 1.7e-08;
Matches 97; Conservative 27; Mismatches 135; Indels 157; Gaps 20;

```

```

DR 235 THG5TQCMCPITYGKDCDTKTKPC----SPSPQNGIGCRSNGLSY---ECKCPKGFBSK 288
QY 176 LCTDCMDGYFSSL-RNETHSICACDESKTCSGLTNRDCGCEYGVWLDGACVYDDEC 234
DB 289 NGEQNTDCLGLHCQNGGCTCIGDISDYTCRCPNFTGRCCQD-----DYDEC 335
QY 235 A-AEPPPSAAOFCNKANGSYCEBCDSCV-GCTG---EPPGCKE-----CISG 280
DB 336 AQRDHPVCNGATCTNTHGSYSY----ICVNGWAGLDCSNNTDCKQAAFCYATCICD 390
QY 281 VAREHGOCA-----IDV 291
DB 391 VGSFTQCTKTKTGLLCHLDADACTSPNHCADALICDTPSNGSYACSCATGYGVDCSEDI 450
QY 292 DECSIAEKTQVRKNENCVTPGSSVYCVCPDGEETEDACVPAAEATGES-PTQ 346
DB 451 DECD--QGSPCEHNGICVNTPGSYRNCSGQFTG-----PCEFNINCESHPQ 498
RESULT 35
FBI5 RAT STANDARD; PRT; 448 AA.
ID FBI5 RAT
AC 09VWH8: 09R284;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Ftblin-5 precursor (Ftbl-5) (Developmental arteries and neural crest
DE EGF-like protein) (Dance) (Embryonic vascular EGF repeat-containing
DE protein) (EVEE).
GN FBI5 OR DANCE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid-10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99357779; PubMed-10428623;
RA Nakamura T., Ruiz-Lozano P., Lindner V., Yabe D., Tanikawa M.,
RA Furukawa Y., Kobuke K., Tashiro K., Lu Z., Andon N.L., Schaub R.,
RA Matsumori A., Sasayama S., Chien K.R., Honjo T.;
RA "DANCE", a novel secreted RGD protein expressed in developing,
RT atherosclerotic, and balloon-injured arteries.";
RT J. Biol. Chem. 274:22476-22483(1999).
RN [2]
RN SEQUENCE FROM N.A.
RA MEDLINE-99278197; PubMed-10347091;
RA Kowal R.C., Richardson J.A., Milano J.M., Olson E.N.;
RA "EVEE", a novel epidermal growth factor-like repeat-containing protein
RT upregulated in embryonic and diseased adult vasculature.";
RT Circ. Res. 84:1166-1176(1999).
RL -I- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
INTERACTION OF INTEGRINS AND THE RGD MOTIF. COULD BE A VASCULAR
LIGAND FOR INTEGRIN RECEPTORS AND MAY PLAY A ROLE IN VASCULAR
DEVELOPMENT AND REMODELING.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation/
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
DB EMBL; AF112153; AAD41769.1; -
DB EMBL; AF137350; AAD25101.1; -
DB HSSP; P00736; IAPQ.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF_Like.
DR InterPro; IPR001881; EGF_CA.
DR Pfam; PF00008; EGF; 4.
DR SMART; SM00179; EGF_CA; 4.

```

[illegible]

	Query Match	10.88;	Score 216;	DB 1;	Length 448;
	Best Local Similarity	29.38;	Pred. No. 3.4e-09;		
	Matches 34; Conservative	12;	Mismatches 52;	Indels 66;	Gaps
OY	217 CEVGMVLDEG-ACVDVECAEAPEPPCSAOPCKNANSTYCE-----	257			
	: ::             : :				
Db	113 CRFGYQMEGNQGVADVEBCATDSHQNCPTQCITITEGYTSCIDGYLWLEGGCLDIDEC	177			
OY	258 ---ECDSVCVGCTGE---GPG-----NKE-----C	277			
	: ::             : :				
Db	173 RYGCQQQLCANVPESYSTCNPFTLNDGRSCQDVNECFENPCVQTCVNTYGSLICRC	233			
OY	278 ISGARARENG--QCADVDECSLAERTCYARKNKNCYNTPGSYCVCPDDGFETED--ACVPP	333			
	: ::             : :				
Db	233 DPGYLELEDGIHCSDMDECSPSEFLC---QHECVNPGSYFCSCPGLVLLDNRSQODI	289			
OY	.334 AEAE 337				
	: ::             : :				
Db	290 NECE 293				
	: ::             : :				
	RESULT 36				
	FBL5_MOUSE				
ID	FBL5_MOUSE	STANDARD:	PRT:	448 AA.	
AC	Q9WVH9;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Fblulin-5 precursor (FBL-5) (Developmental arteries and neural crest				
DE	Egr-like protein) (Dance).				
GN	FBLN5 OR DANCE.				
OS	Mus musculus (Mouse).				
OC	Euryarchaea; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				

RX	MEDLINE-9935779; PubMed=10428823;
RA	Nakamura T., Ruiz-Lozano P., Lindner V., Yabe D., Tanitaki M.,
RA	Furukawa Y., Kozube K., Tashiro K., Lu Z., Andon N.L., Schaud R.,
RA	Matsuura A., Sasayama S., Chen K.R., Honjo T.;
RT	"DANCE", a novel secreted RGD protein expressed in developing,
RT	atherosclerotic, and balloon-injured arteries.";
RL	J. Biol. Chem. 274:22476-22483(1999).
CC	-I- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
CC	INTERACTION OF INTEGRINS AND THE RGD MOTIF. COULD BE A VASCULAR
CC	LIGAND FOR INTERBETA RECEPTORS AND MAY PLAY A ROLE IN VASCULAR
CC	DEVELOPMENT AND REMODELING.
CC	-I- SUBCELLULAR LOCATION: Secreted.
CC	-I- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AF112151; AAP41767.1; "
DR	HSSP; P00736; IAPQ.
DR	MGP; MG1:1346091; Fbln5.
DR	InterPro; IPRO00152; Asx_hydroxy1.
DR	InterPro; IPRO00561; EGF-like.
DR	InterPro; IPRO01881; EGF_Ca.
DR	Pfam; PF00008; EGF; 4.
DR	SMART; SMART0179; EGF_Like; 4.
DR	SMART; SM00001; EGF_Like; 2.
DR	PROSITE; PS00010; ASX-HYDROXY; 4.
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.
DR	PROSITE; PS01186; EGF_2; 4.
KW	PROSITE; PS01187; EGF_CA; 6.
DR	Repeat: EGF-like domain; Calcium-binding; Glycoprotein; Signal.
FT	SIGNAL
FT	CHAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	SITE
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	DISULFID
FT	CARBOHYD
FT	CARBOHYD
FO	SEQUENCE
	448 AA; 50193 MW; F15CCK7OCCEFC97 CRG64;
	N-LINKED (GLCNAC... ) (POTENTIAL).
	N-LINKED (GLCNAC... ) (POTENTIAL).
	F15CCK7OCCEFC97 CRG64;

```

Query Match 10.7% Score 214.5 DB 1 length 448:
Best Local Similarity 36.2% Pred. No. 4.3e-09:
Matches 47; Conservative 7; Mismatches 37; Indels 39; Gaps 5

OY 217 CEVGNVLDGE-ACVYDDECAEPPPCSAQPFKNANGSTCEECDSVCYGTGEGPNCCK 275
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 113 CRFGYMGEGNCVCVYDDECATSHOCNPPQIDINTEGGYTC----- 153

```

Oy	276	ECISSTAREHGCADVDCSCSLAEKTCYRKNNGCNFTGSGYCVC PDGE-----		320
Db	154	SCTDGYWLLBEGCCLDIDBCRG--YC---QQLCANVFQSYICTGNPGFTLNDGRSQDV		208
Oy	325	--ETEDACV	331	
Db	209	NCECFENPCV	218	
RESULT	37			
NIDO_MOUSE		STANDARD;	PRT; 1245 AA.	
ID	NIDO_MOUSE			
AC	p10493:			
Df	01-APR-1990 (Rel. 14, Created)			
Dt	01-APR-1990 (Rel. 14, Last sequence update)			
Df	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Nitrogen precursor (Entactin).			
GN	NID OR NIDI OR ENT.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCL_TaxID=10990;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 29-40.			
RX	MEDLINE=89079780; PubMed=3264556;			
RA	Durkin M.E., Chakravarti S., Bartos B.B., Liu S.H., Friedman R.L.,			
RT	Chung A.E.;			
RT	"Amino acid sequence and domain structure of entactin. Homology with			
RT	epidermal growth factor precursor and low density lipoprotein			
RT	receptor.";			
RL	J. Cell Biol. 107:2749-2756(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=89231638; PubMed=2496973;			
RA	Mann K., Deutzmann R., Annaileij M., Timpl R., Ramondi L.,			
RA	Yamada Y., Pan T.-C., Conway D., Chu M.-I.;			
RT	"Amino acid sequence of mouse nidogen, a multidomain basement			
RT	membrane protein with binding activity for laminin, collagen IV and			
RT	cells.";			
RL	EMBO J. 8:65-72(1989).			
RN	[3]			
RP	SEQUENCE OF 1-251 FROM N.A.			
RC	STRAIN=BALB/C; TISSUE=Liver;			
RX	MEDLINE=94040771; PubMed=8224873;			
RA	Durkin M.E., Liu S.H., Reing J., Chung A.E.;			
RT	"Characterization of the 5' end of the mouse Ent gene encoding the			
RT	basement membrane protein, entactin.";			
RL	Gene 132:261-266(1993).			
RN	[4]			
RP	SEQUENCE OF 1207-1245 FROM N.A.			
RC	STRAIN=C57BL/6J X CBA/J;			
RX	MEDLINE=95324912; PubMed=7601446;			
RA	Durkin M.E., Wewer U.M., Chung A.E.;			
RT	"Exon organization of the mouse entactin gene corresponds to the			
RT	structural domains of the polypeptide and has regional homology to			
RT	the low-density lipoprotein receptor gene.";			
RL	Genomics 26:219-228(1995).			
RN	[5]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE=86192477; PubMed=3084254;			
RA	Paulsson M., Deutzmann R., Dziadek M., Nowack H., Timpl R., Weber S.,			
RA	Engel J.;			
RT	"Purification and structural characterization of intact and			
RT	fragmented nidogen obtained from a tumor basement membrane.";			
RL	Bur. J. Biochem. 156:467-478(1986).			
RN	[6]			
RP	CARBOHYDRATE-LINKAGE SITES, AND PARTIAL SEQUENCE.			
RX	MEDLINE=93316903; PubMed=8326911;			
RA	Fujiwara S., Shinkai H., Mann K., Timpl R.;			
RT	"Structure and localization of O- and N-linked oligosaccharide chains			
RT	on basement membrane protein nidogen.";			
RL	Matrix 13:215-222(1993).			
CC	-1- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN			

CC	BASMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.
CC	ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-
CC	EXTRACELLULAR MATRIX INTERACTIONS.
CC	-I- SUBCELLULAR LOCATION: BASMENT MEMBRANES.
CC	-I- PTM: N- AND O-GLYCOSYLATED.
CC	-I- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC	-I- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC	-I- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X14194; CAA32408.1; -;
DR	EMBL; X14480; CAA32642.1; -;
DR	EMBL; L117324; AAA7652.1; -;
DR	EMBL; L117322; AAA7652.1; JOINED.
DR	EMBL; L117323; AAA7652.1; JOINED.
DR	EMBL; X83093; CAA58148.1; -;
DR	PIR; S02730; MMSND.
DR	HSSP; P07204; IADX.
DR	CarbBank; CCSD:37018; -;
DR	CarbBank; CCSD:37019; -;
DR	MGP; MGI:97342; Nid1.
DR	InterPro; IPRO000152; Asx_hydroxyl.
DR	InterPro; IPRO000561; EGF-like.
DR	InterPro; IPRO01881; EGF-Ca.
DR	InterPro; IPRO000033; Idl_receptor_rep.
DR	InterPro; IPRO03886; Nidogen_ext.
DR	InterPro; IPRO00716; Thyroglobulin_1.
DR	pfam; PF00008; EGF; 6.
DR	pfam; PF00058; Idl_recept_b; 3.
DR	pfam; PF00085; thyroglobulin_1; 1.
DR	SMART; SM00179; EGF_CA; 2.
DR	SMART; SM00001; EGF_Like; 4.
DR	SMART; SM00135; LY; 5.
DR	SMART; SM00539; NIDO; 1.
DR	SMART; SMG0211; TY; 1.
DR	PROSITE; PS00010; ASX_HYDROXYL; 3.
DR	PROSITE; PS00022; EGF_1; 1.
DR	PROSITE; PS00484; THYROGLOBULIN_1; 1.
DR	PROSITE; PS01186; EGF_2; 4.
DR	PROSITE; PS01187; EGF_CA; 2.
KV	Basement membrane: Extracellular matrix; Glycoprotein; Sulfation;
KW	Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.
FT	SIGNAL 1 28
FT	CHAIN 29 1245 NIDGEN
FT	DOMAIN 29 667 I (LARGER GLOBULAR DOMAIN).
FT	DOMAIN 668 919 II (CYSTEINE-RICH).
FT	DOMAIN 920 1245 III (SMALLER GLOBULAR DOMAIN).
FT	DOMAIN 384 424 EGF-Like 1.
FT	DOMAIN 666 707 EGF-Like 2.
FT	DOMAIN 708 749 EGF-Like 3, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 756 799 EGF-Like 4.
FT	DOMAIN 800 838 EGF-Like 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 870 917 THYROGLOBULIN TYPE 1.
FT	DOMAIN 987 1028 LDL-RECEPTOR YWTD MOTIF 1.
FT	DOMAIN 1030 1071 LDL-RECEPTOR YWTD MOTIF 2.
FT	DOMAIN 1073 1116 LDL-RECEPTOR YWTD MOTIF 3.
FT	DOMAIN 1122 1161 LDL-RECEPTOR YWTD MOTIF 4.
FT	DOMAIN 1206 1242 EGF-Like 6.
FT	CAROAHN 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CAROAHN 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MOD_RES 290 290 SULFATION (POTENTIAL).
FT	MOD_RES 295 295 SULFATION (POTENTIAL).
FT	DISULFD 670 683 BY SIMILARITY.
FT	DISULFD 677 693 BY SIMILARITY.
FT	DISULFD 695 706 BY SIMILARITY.
FT	DISULFD 712 725 BY SIMILARITY.



Query Match	10.6%;	Score 213.5;	DB 1;	Length 1245;
Best Local Similarity	27.5%;	Pred. No. 1.3e-08;		
Matches 53; Conservative	27;	Mismatches 70;	Indels 43;	Gaps 8;

RESULT	38	
NTC3_MOUSE		
NTC3_MOUSE		
STANDARD;		
PRT;	2318	AA

CC	-1- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.	
CC	-1- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.	
CC	-1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.	
CC	-1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.	
CC	-1- SIMILARITY: CONTAINS 5 ANK REPEATS.	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL: X74760; CAAS2776.1; -	
DR	HSSP: P00740; IIXA.	
DR	MGD: MGI:99460; Notch3.	
DR	InterPro: IPR002110; ANK.	
DR	InterPro: IPR000152; Asx_hydroxyl.	
DR	InterPro: IPR000561; EGF-Like.	
DR	InterPro: IPR000742; EGF_2.	
DR	InterPro: IPR001881; EGF_Ca.	
DR	InterPro: IPR001438; EGF_II.	
DR	InterPro: IPR000800; Notch.	
DR	Pfam: PF00023; ank; 6.	
DR	Pfam: PF00008; EGF; 34.	
DR	Pfam: PF00066; notch; 3.	
DR	PRINTS: PR00010; EGFBLD.	
DR	PRINTS: PR01452; NOTCH.	
DR	SMART: SM00248; ANK; 5.	
DR	SMART: SM00119; EGF_Ca; 19.	
DR	SMART: SM00001; EGF_Like; 13.	
DR	SMART: SM00004; NL; 3.	
DR	PROSITE: PS50088; ANK_REPEAT; 4.	
DR	PROSITE: PS50297; ANK_REP_REGION; 1.	
DR	PROSITE: PS00010; ASX_HYDROXYL; 18.	
DR	PROSITE: PS00022; EGF_1; 33.	
DR	PROSITE: PS01186; EGF_2; 27.	
DR	PROSITE: PS01187; EGF_Ca; 17.	
KW	Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;	
KW	ANK repeat; Glycoprotein.	
FT	DOMAIN	1 1643
FT	TRANSMEM	1644 1664
FT	DOMAIN	1665 2318
FT	DOMAIN	2242 2261
FT	DOMAIN	39 78
FT	DOMAIN	79 119
FT	DOMAIN	120 157
FT	DOMAIN	159 196
FT	DOMAIN	198 235
FT	DOMAIN	237 273
FT	DOMAIN	275 313
FT	DOMAIN	315 351
FT	DOMAIN	352 390
FT	DOMAIN	392 430
FT	DOMAIN	432 468
FT	DOMAIN	470 506
FT	DOMAIN	508 544
FT	DOMAIN	546 581
FT	DOMAIN	583 619
FT	DOMAIN	619 656
FT	DOMAIN	658 694
FT	DOMAIN	696 731
FT	DOMAIN	735 771
FT	DOMAIN	772 809
FT	DOMAIN	811 848
FT	DOMAIN	850 886
FT	DOMAIN	888 923
FT	DOMAIN	925 961
FT	DOMAIN	963 999
FT	DOMAIN	1001 1035
FT	DOMAIN	1037 1083
FT	DOMAIN	1085 1121

```

FT DOMAIN 1123 1159 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1161 1204 EGF-Like 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1204 1245 EGF-Like 31.
FT DOMAIN 1247 1288 EGF-Like 32.
FT DOMAIN 1290 1326 EGF-Like 33.
FT DOMAIN 1336 1374 EGF-Like 34.
FT REPEAT 1388 1428 LIN/NOTCH 1.
FT REPEAT 1429 1467 LIN/NOTCH 2.
FT REPEAT 1468 1503 LIN/NOTCH 3.
FT REPEAT 1839 1868 ANK 1.
FT REPEAT 1872 1902 ANK 2.
FT REPEAT 1906 1935 ANK 3.
FT REPEAT 1939 1968 ANK 4.
FT REPEAT 1972 2001 ANK 5.
FT DISULFID 43 55 BY SIMILARITY.
FT DISULFID 49 66 BY SIMILARITY.
FT DISULFID 68 77 BY SIMILARITY.
FT DISULFID 83 94 BY SIMILARITY.
FT DISULFID 88 107 BY SIMILARITY.
FT DISULFID 109 118 BY SIMILARITY.
FT DISULFID 124 135 BY SIMILARITY.
FT DISULFID 129 145 BY SIMILARITY.
FT DISULFID 147 156 BY SIMILARITY.
FT DISULFID 163 175 BY SIMILARITY.
FT DISULFID 169 184 BY SIMILARITY.
FT DISULFID 186 195 BY SIMILARITY.
FT DISULFID 202 213 BY SIMILARITY.
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 225 234 BY SIMILARITY.
FT DISULFID 241 252 BY SIMILARITY.
FT DISULFID 246 261 BY SIMILARITY.
FT DISULFID 263 272 BY SIMILARITY.
FT DISULFID 279 292 BY SIMILARITY.
FT DISULFID 286 301 BY SIMILARITY.
FT DISULFID 303 312 BY SIMILARITY.
FT DISULFID 319 330 BY SIMILARITY.
FT DISULFID 324 339 BY SIMILARITY.
FT DISULFID 341 350 BY SIMILARITY.
FT DISULFID 356 367 BY SIMILARITY.
FT DISULFID 361 378 BY SIMILARITY.
FT DISULFID 380 389 BY SIMILARITY.
FT DISULFID 386 409 BY SIMILARITY.
FT DISULFID 403 418 BY SIMILARITY.
FT DISULFID 420 429 BY SIMILARITY.
FT DISULFID 436 447 BY SIMILARITY.
FT DISULFID 441 456 BY SIMILARITY.
FT DISULFID 458 467 BY SIMILARITY.
FT DISULFID 474 485 BY SIMILARITY.
FT DISULFID 479 494 BY SIMILARITY.
FT DISULFID 496 505 BY SIMILARITY.
FT DISULFID 512 523 BY SIMILARITY.
FT DISULFID 517 532 BY SIMILARITY.
FT DISULFID 534 543 BY SIMILARITY.
FT DISULFID 543 560 BY SIMILARITY.
FT DISULFID 550 569 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 587 598 BY SIMILARITY.
FT DISULFID 592 607 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT DISULFID 625 635 BY SIMILARITY.
FT DISULFID 630 644 BY SIMILARITY.
FT DISULFID 646 655 BY SIMILARITY.
FT DISULFID 662 673 BY SIMILARITY.
FT DISULFID 667 682 BY SIMILARITY.
FT DISULFID 684 693 BY SIMILARITY.
FT DISULFID 700 710 BY SIMILARITY.
FT DISULFID 705 719 BY SIMILARITY.
FT DISULFID 721 730 BY SIMILARITY.
FT DISULFID 739 750 BY SIMILARITY.
FT DISULFID 744 759 BY SIMILARITY.
FT DISULFID 761 770 BY SIMILARITY.
FT DISULFID 776 787 BY SIMILARITY.
FT DISULFID 781 797 BY SIMILARITY.

```

```

FT DISULFID 799 808 BY SIMILARITY.
FT DISULFID 815 827 BY SIMILARITY.
FT DISULFID 821 836 BY SIMILARITY.
FT DISULFID 838 847 BY SIMILARITY.
FT DISULFID 854 865 BY SIMILARITY.
FT DISULFID 859 874 BY SIMILARITY.
FT DISULFID 876 885 BY SIMILARITY.
FT DISULFID 892 902 BY SIMILARITY.
FT DISULFID 897 911 BY SIMILARITY.
FT DISULFID 913 922 BY SIMILARITY.
FT DISULFID 929 940 BY SIMILARITY.
FT DISULFID 934 949 BY SIMILARITY.
FT DISULFID 951 960 BY SIMILARITY.
FT DISULFID 967 978 BY SIMILARITY.
FT DISULFID 972 987 BY SIMILARITY.
FT DISULFID 989 998 BY SIMILARITY.
FT DISULFID 1005 1016 BY SIMILARITY.
FT DISULFID 1010 1023 BY SIMILARITY.
FT DISULFID 1025 1034 BY SIMILARITY.
FT DISULFID 1041 1062 BY SIMILARITY.
FT DISULFID 1056 1071 BY SIMILARITY.
FT DISULFID 1073 1082 BY SIMILARITY.
FT DISULFID 1089 1100 BY SIMILARITY.
FT DISULFID 1094 1109 BY SIMILARITY.
FT DISULFID 1111 1120 BY SIMILARITY.
FT DISULFID 1127 1138 BY SIMILARITY.
FT DISULFID 1132 1147 BY SIMILARITY.
FT DISULFID 1149 1158 BY SIMILARITY.
FT DISULFID 1165 1183 BY SIMILARITY.
FT DISULFID 1177 1192 BY SIMILARITY.
FT DISULFID 1194 1203 BY SIMILARITY.

Query Match 10.6%; Score 212; DB 1; Length 2318;
Best Local Similarity 28.2%; Pred. No. 3.1e-08;
Matches 73; Conservative 30; Mismatches 86; Indels 70; Gaps 18;

QY 128 CSBGTGYPCL--ACQSGSRPCSGNGHSGD---GSGQSGSCRCNMGYGLCTD-- 179
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 68 CLFGWGEBCOLEDPCHSG--PCAGRGVCOSSVAVGTAR--FSCRCLEGFQGPDCSQPD 122
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 180 -----CMGGRSSLNENHSTICTAC-----DESCRT-----C-SGLTNRDCG----- 215
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 123 PCVSRPCVHGAPOCVGPDPRFAC-ACPPGYGQSCQSDIDECSRSGTCHRGCTCNTPGS 181
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 216 ---ECEVGV--VLDEGACVVDDECAEPPCSAOPCKNANG--SYTC-----ECCDS 261
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 FRQGPGLGTGLGLENPVY---PCA--PSPCRNNGTGRQSSDYTYDCACLPBEGQNCYEV 236
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 SCVGTGCEGPCKCKECISGYAREHGQC-----ADVDCSLAEKTCVRKNENCYNT 311
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 237 NVDDCPGHCLNGCTGVGVNVTYNCQCPPEWTGQFCTEDVDECQLQPNAC--HNGGTGFNL 295
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 312 PGSYVVCVCPDGFPEPFDAC 330
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 296 LGHSHSCVANGW--TGESC 312
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 39
NEL1_HUMAN STANDARD; PRT; 810 AA.
ID NEL1_HUMAN Q9Y472;
AC Q9Y472;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein kinase C-binding protein NEL1 precursor (NEL1-like protein 1)
DE (Nel-related protein 1).
GN NEL1 OR NRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

CC TISSUE-Brain.  
RX MEDLINE-97131504; PubMed-8975702;  
RA Matanabe T.K., Katagiri T., Suzuki M., Shimizu F., Fujiwara T.,  
RA Kanemoto N., Nakamura Y., Hirai Y., Maekawa H., Takahashi E.,  
RT "Cloning and characterization of two novel human CDNs (NELL1 and  
NELL2) encoding proteins with six EGF-like repeats."  
RL Genomics 38:273-276(1996).  
RN [2]  
RP SEQUENCE OF 383-810 FROM N.A.  
RA Ting K., Vastardis H., Mulliken J.B., Bertolami C., Wen Z.,  
RA Young M., Tleu A., Kwong E.,  
RT "Nel homolog gene expression in craniofacial anomalies."  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
CC -1 SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1 (BY SIMILARITY).  
CC -1 CELLULAR LOCATION: Secreted (By similarity).  
CC -1 DISEASE: EXPRESSED IN CRANIOFACIAL ANOMALIES.  
CC -1 SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.  
CC -1 SIMILARITY: CONTAINS 5 WMFC DOMAINS.  
CC -1 SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
CC -1 CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
IN POSITIONS 427 AND 771.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D83017; BAAL1680.1; -  
DR EMBL; U57523; AAB06946.1; ALT\_FRAME.  
DR HSSP; P07204; IADX.  
DR MIM; 602319; -  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001881; EGF\_CA.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR003129; TSPN.  
DR InterPro; IPR001007; WMFC.  
DR Pfam; PF00008; EGF; 4.  
DR Pfam; PF00093; vwc; 2.  
DR SMART; SM00179; EGF\_CA; 2.  
DR SMART; SM00001; EGF-like; 4.  
DR SMART; SM00282; Lamg; 1.  
DR SMART; SM00210; TSPN; 1.  
DR SMART; SM00214; vwc; 3.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 3.  
DR PROSITE; PS01187; EGF\_CA; 3.  
DR PROSITE; PS01208; WMFC; 2.  
KW Glycoprotein; EGF-like domain; Repeat; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 810  
FT DOMAIN 81 230  
FT DOMAIN 230 331  
FT DOMAIN 335 390  
FT DOMAIN 391 433  
FT DOMAIN 434 475  
FT DOMAIN 476 516  
FT DOMAIN 515 547  
FT DOMAIN 549 595  
FT DOMAIN 596 631  
FT DOMAIN 632 687  
FT DOMAIN 687 750  
FT DOMAIN 752 807  
FT DOMAIN 807 846  
FT DISULFID 401 416  
FT DISULFID 418 432  
FT DISULFID 438 451  
FT DISULFID 445 460  
FT DISULFID 462 474

FT DISULFID 480 493 BY SIMILARITY.  
FT DISULFID 487 502 BY SIMILARITY.  
FT DISULFID 504 515 BY SIMILARITY.  
FT DISULFID 519 529 BY SIMILARITY.  
FT DISULFID 523 535 BY SIMILARITY.  
FT DISULFID 537 546 BY SIMILARITY.  
FT DISULFID 553 566 BY SIMILARITY.  
FT DISULFID 560 575 BY SIMILARITY.  
FT DISULFID 577 594 BY SIMILARITY.  
FT DISULFID 600 613 BY SIMILARITY.  
FT DISULFID 607 622 BY SIMILARITY.  
FT DISULFID 624 630 BY SIMILARITY.  
FT DISULFID 640 640 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 53 53 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 83 83 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 224 224 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 294 294 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 372 372 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 511 511 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 562 562 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 609 609 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 708 708 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 732 732 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 758 758 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CONFLICT 383 383 N -> D (IN REF. 2).  
FT CONFLICT 573 573 Y -> H (IN REF. 2).  
FT CONFLICT 626 626 S -> C (IN REF. 2).  
SQ SEQUENCE 810 AA: 89606 MM: 549465EAF7AEED0 CRC64:

Query Match 10.5%; Score 211.5; DB 1; Length 810;  
Best Local Similarity 22.0%; Pred. No. 1.3e-08;  
Matches 95; Conservative 30; Mismatches 122; Indels 185; Gaps 21.

OY 28 PTPCHRCRLVDKFN--QGMVD-----TAKNNGGNTAMEKTSKYSSSEIRLE 77  
DB 221 PNLNHTCPTCSDFLSLVQGIIMDLQELAKMTAKLYVA-----ETRLSOLENCHEKTC 273  
OY 78 ILEGL-----CESSDPECNOMLEAGEEHLQAWMLQKSEYPLLEWF 119  
DB 274 QVSGLLYRDQDSWDGDHCRCNCTCKSGAVECRM----- 307  
OY 120 CVKTLKVCSPPTGYRDCIACGGSGQ-----NP-CSGNHCSGDSRQDSCRCRMGYQ 173  
DB 308 -----SCPRLNCSRPSDLPIVHAGCCVCKVCPKCIYGVKVLAEQRILTKSCR----- 354  
OY 174 GPLCTDCMDGYSSLRNETHSI-CTACD-----ESKTCGSLT-----NRDCG- 215  
DB 355 -----ECRGGVLYKITEMCPRPLNCKSEKHILPENOCRCVCRGHNFCABEGPKCGENSECKN 409  
OY 216 -----RCVEGVWVL-----DEGACVDVDECAARPPCSAOPCKNNNGSYTCECCSSCV- 264  
DB 410 WNTFKATCECKSGYSIVQGSAYCEDIDECAAMHYCHANTVAVNLPGLYRCD-----CVP 464  
OY 265 -----GCT-----GEGPGNCKE-----CISGY-----AR 283  
DB 465 GYIRVDPSCTEHDECGSQHNCNDENALCTNTVQGHSCRCRKGYYGNGTICARFCEEGCR 524  
OY 284 EHGOCA-----DVDECSLAKEKTCVRKNNKNCYNTPGSYCVCPDGEFE 325  
DB 525 YGCTVAPARKVCYCPSGFTGSHCEKIDIDECSEGIIEC-HNHSRCVNLPGWYHCECRSGFHD 583  
OY 326 -----TEDACV 331  
DB 584 DGTYSLSGESCT 595  
RESULT 40  
EMBL\_MOUSE STANDARD: PRT: 931 AA.  
ID EMBL\_MOUSE  
AC 061549;  
DT 01-NOV-1997 (rel. 35, Created)  
DT 01-NOV-1997 (rel. 35, Last sequence update)

DT 01-MAR-2002 (Rel. 41, last annotation update)  
DE Cell surface glycoprotein EMRI precursor (EMRI hormone receptor)  
GN EMRI OR GPF480.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BAIB/C; TISSUE=Peritoneal cavity;  
RX MEDLINE=96132946; PubMed=8550607;  
RA McKnight A.J., Macfarlane A.J., Dri P., Turley L., Willis A.C.,  
RA Gordon S.;  
RT "Molecular cloning of F4/80, a murine macrophage-restricted cell  
RT surface glycoprotein with homology to the G-protein-linked  
RT transmembrane 7 hormone receptor family.";  
RL J. Biol. Chem. 271:486-489(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97312684; PubMed=9169125;  
RA Lin H.H., Stubbs L.J., Mucenski M.L.;  
RT "Identification and characterization of a seven transmembrane hormone  
RT receptor using differential display.";  
RL Genomics 41:301-308(1997).  
CC -1- FUNCTION: PROBABLY INVOLVED IN CELL ADHESION WITHIN TISSUES  
CC AND RECEPTOR SIGNALING.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- TISSUE SPECIFICITY: IN MACROPHAGES, BUT ABSENT FROM THOSE WHICH  
CC ARE LOCALIZED WITHIN T-CELL AREAS OF LYMPH NODES AND SPLEEN.  
CC LOW LEVEL OF EXPRESSION ON BLOOD MONOCYTES.  
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC  
CC EMBL: X93328; CAA63720.1; -  
DR EMBL: 066888; AAC53184.1; -  
DR HSSP: P07204; 1FGD.  
DR GCRDB: GCR\_1309; -  
DR MGD: MGI:106912; Emr1.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR000832; GPCR\_secretin.  
DR InterPro: IPR000203; PKD\_cys\_rich.  
DR Pfam: PF00002; 7tm\_2; 1.  
DR Pfam: PF00008; EGF; 6.  
DR Pfam: PF01825; GPS; 1.  
DR PRINTS: PRO1128; EMRIHORMONER.  
DR SMART: SM00179; EGF\_CA; 6.  
DR SMART: SM00001; EGF\_Like; 1.  
DR SMART: SM00303; GPS; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 6.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS01187; EGF\_CA; 5.  
DR PROSITE: PS00221; GPS; 1.  
DR PROSITE: PS00650; G\_PROTEIN\_RECP\_F2\_2; 1.  
DR PROSITE: PS00261; G\_PROTEIN\_RECP\_F2\_4; 1.  
DR G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;  
KW EGF-like domain; Repeat; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 931  
FT DOMAIN 28 644  
FT TRANSMEM 645 672  
FT DOMAIN 673 679  
FT POTENTIAL.  
FT CELL SURFACE GLYCOPROTEIN EMRI.  
FT EXTRACELLULAR (POTENTIAL).  
FT POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 680 701  
FT DOMAIN 702 711  
FT TRANSMEM 712 735  
FT DOMAIN 736 754  
FT TRANSMEM 755 776  
FT DOMAIN 777 792  
FT TRANSMEM 793 821  
FT DOMAIN 822 839  
FT TRANSMEM 840 859  
FT DOMAIN 860 874  
FT TRANSMEM 875 897  
FT DOMAIN 898 931  
FT TRANSMEM 931 931  
FT DOMAIN 81 132  
FT DOMAIN 133 172  
FT DOMAIN 173 221  
FT DOMAIN 222 271  
FT DOMAIN 272 318  
FT DOMAIN 319 367  
FT DOMAIN 367 415  
FT SITE 506 508  
FT DISULFID 36 48  
FT DISULFID 42 57  
FT DISULFID 59 79  
FT DISULFID 85 98  
FT DISULFID 92 107  
FT DISULFID 109 131  
FT DISULFID 137 149  
FT DISULFID 143 158  
FT DISULFID 160 171  
FT DISULFID 177 189  
FT DISULFID 183 198  
FT DISULFID 200 220  
FT DISULFID 226 239  
FT DISULFID 233 248  
FT DISULFID 250 270  
FT DISULFID 276 286  
FT DISULFID 280 295  
FT DISULFID 297 317  
FT DISULFID 323 336  
FT DISULFID 330 345  
FT DISULFID 347 366  
FT CARBOHYD 148 167  
FT CARBOHYD 167 187  
FT CARBOHYD 229 239  
FT CARBOHYD 259 269  
FT CARBOHYD 283 283  
FT CARBOHYD 405 405  
FT CARBOHYD 417 417  
FT CARBOHYD 474 474  
FT CARBOHYD 498 498  
FT CARBOHYD 706 706  
FT SEQUENCE 931 AA; 102129 MM; 529634667E8876B5 CMC64;  
SO  
Query Match 10.5%; Score 211.5; DB 1; Length 931;  
Best Local Similarity 28.2%; Pred. No. 1.4e-08;  
Matches 64; Conservative 19; Mismatches 77; Indels 67; Gaps 12;  
OY 148 CSNGHCSGSDGSRGSGSCRCMHG--QGPLETD---CMDGTFSSLRNETHSIC---TVC 199  
DB 143 CPKYSNCS---NSVGSYSCTCGPGFVLNGLSICEDDEDEC-----TRVCPENHATC 189  
OY 200 DES-----CKTSGGLNRRCCGECEVGWVDEGACVVDCAEAPPCSAAGCKNANGSY 254  
DB 190 HNTLGSYCTCNSGLSSGGGPPFGG--LDF--SCEDVDGCSRNSITLGGFTFICINTLGSY 246  
OY 255 TC-----ECCDSSC---VGCSTGEGPNCCKECISGVAREHG 286  
DB 247 SCSCPAGFSLPFPQILGHADGNCITDIDCDPTCPNLSCTWTGYSFCTCHGFASSNG 306  
OY 287 Q-----CADVDECSIAERTCVAKNENCNTPGSIYCVCPDGE 324



```

FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 298 311 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 380 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
FT DISULFID 415 428 BY SIMILARITY.
FT DISULFID 422 437 BY SIMILARITY.
FT DISULFID 439 448 BY SIMILARITY.
FT DISULFID 455 466 BY SIMILARITY.
FT DISULFID 460 475 BY SIMILARITY.
FT DISULFID 477 486 BY SIMILARITY.
FT DISULFID 493 503 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 568 578 BY SIMILARITY.
FT DISULFID 573 587 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 605 616 BY SIMILARITY.
FT DISULFID 610 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 648 662 BY SIMILARITY.
FT DISULFID 664 673 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 685 700 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 718 728 BY SIMILARITY.
FT DISULFID 723 737 BY SIMILARITY.
FT DISULFID 739 748 BY SIMILARITY.
FT DISULFID 755 766 BY SIMILARITY.
FT DISULFID 760 775 BY SIMILARITY.
FT DISULFID 777 786 BY SIMILARITY.
FT DISULFID 793 804 BY SIMILARITY.
FT DISULFID 798 813 BY SIMILARITY.
FT DISULFID 815 824 BY SIMILARITY.
FT DISULFID 831 842 BY SIMILARITY.
FT DISULFID 836 853 BY SIMILARITY.
FT DISULFID 855 864 BY SIMILARITY.
FT DISULFID 871 882 BY SIMILARITY.
FT DISULFID 876 891 BY SIMILARITY.
FT DISULFID 893 902 BY SIMILARITY.
FT DISULFID 909 920 BY SIMILARITY.
FT DISULFID 914 929 BY SIMILARITY.
FT DISULFID 931 940 BY SIMILARITY.
FT DISULFID 947 958 BY SIMILARITY.
FT DISULFID 952 967 BY SIMILARITY.
FT DISULFID 969 978 BY SIMILARITY.
FT DISULFID 1023 1034 BY SIMILARITY.
FT DISULFID 1028 1043 BY SIMILARITY.
FT DISULFID 1045 1054 BY SIMILARITY.

```

Query Match 10.5%; Score 211.5; DB 1; Length 2437;  
 Best Local Similarity 20.7%; Pred. No. 3.5e-08;  
 Matches 71; Conservative 21; Mismatches 118; Indels 133; Gaps 13;

```

QY 128 CPGGTGPPDLACGGSGQRPCSGNGHCS-----GDGSRQ----- 161
DB 128 CPGGSGKTCQLADPCASNCPANGGOCSAFESHYICTCPNPHGQTCRODVNCCAVSPSP 187
QY 162 -----GDGSCRHMGYOGPLC-----TDCMDG 183
DB 188 CRNGGTCINVGSYLCRCPPRYTGPHCQRLXQPLSPCRSGGTGVQTSITHTTSCSLRG 247
QY 184 YFSSLRNETHSICT--ACDESKTCGSLTNRDCGECGEVGNWIDEGACVDVDECAAEPPPC 241

```

```

DB 248 FTGQTCENHVVDDCTOHACENGGBPCIDIGINTYNC-HCDKHMT-GQYCTEDVDECELSPNAG 305
QY 242 SAAQFCNKANGSYTC-----EECDSS---CVGCTGEGPGNCKE----- 276
DB 306 QNCGTCHNHTIGFHCVCVNGWGTGDDCSENIIDDCASACSHGATCDRVASFCECPHGT 365
QY 277 -----CISQYAREHQOC-----ADVDECSLAERTCYVR 304
DB 366 GLLCHDDACISNPGCKSGNSCDTNPVSGKAICTCPPGYTSACNDIDECISGANPC-EH 424
QY 305 NENGYTPSGSYVCVCPDGEFEETEDACVPPAEATGES-PIQ 346
DB 425 GGRCLNTKGSFOCKLQGYEG-----PRCEMDVNCCKSNPCQ 461

RESULT 42
FBL5_HUMAN
ID FBL5_HUMAN STANDARD; PRT; 448 AA.
AC Q9UBX5; O75966;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibulin-5 precursor (FBL-5) (developmental arteries and neural crest
  EGF-like protein) (Dance) (Urine p50 protein) (UP50).
GN FBLN5 OR DANCE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RA Kostka G.;
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99357779; Pubmed=10428823;
RA Nakamura T., Ruiz-Iozano P., Lindner V., Yabe D., Taniwaki M.,
  Furukawa Y., Kobuke K., Tashiro K., Lu Z., Andon N.L., Schaub R.,
  Matsumori A., Sasaizawa S., Chien K.R., Honjo T.;
  "DANCE, a novel secreted RGD protein expressed in developing,
  atherosclerotic, and balloon-injured arteries."
  J. Biol. Chem. 274:22476-22483(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=urine;
RA Zemel R., Sholto O., Shaul Y.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
  INTERACTION OF INTEGRINS AND THE RGD MOTIF. COULD BE A VASCULAR
  LIGAND FOR INTEGRIN RECEPTORS AND MAY PLAY A ROLE IN VASCULAR
  DEVELOPMENT AND REMODELING.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN HEART, OVARY, AND
  COLON BUT ALSO IN KIDNEY, PANCREAS, TESTIS, LUNG, AND PLACENTA.
  NOT DETECTABLE IN BRAIN, LIVER, THYMUS, PROSTATE, OR PERIPHERAL
  BLOOD LEUCOCYTES.
CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation:
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sdb.ch/announce/
  or send an email to licenselsdb-sdb.ch).
CC
CC EMBL: AJ133490; CAB38568.1; -
CC EMBL: AF112152; AAD41768.1; -
CC EMBL: AF093118; AAC62107.1; -
CC HSP: P07204; IFGD.
CC MIM: 604580; -.

```

DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR Pfam: PFO0008; EGF\_4.  
 DR SMART: SM00179; EGF\_Ca; 4.  
 DR SMART: SM00001; EGF\_Like; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 4.  
 DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE: PS01186; EGF\_2; 4.  
 DR PROSITE: PS01187; EGF\_Ca; 6.  
 KW Repeat: EGF-like domain; Calcium-binding; Glycoprotein; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 1 23  
 FT DOMAIN 24 448  
 FT DOMAIN 127 69  
 FT DOMAIN 167 167  
 FT DOMAIN 168 206  
 FT DOMAIN 207 246  
 FT DOMAIN 247 287  
 FT DOMAIN 288 333  
 FT SITE 54 56  
 FT DISULFID 131 144  
 FT DISULFID 138 153  
 FT DISULFID 155 166  
 FT DISULFID 172 181  
 FT DISULFID 177 190  
 FT DISULFID 192 205  
 FT DISULFID 211 221  
 FT DISULFID 217 230  
 FT DISULFID 232 245  
 FT DISULFID 251 262  
 FT DISULFID 258 271  
 FT DISULFID 273 286  
 FT DISULFID 292 305  
 FT DISULFID 299 314  
 FT DISULFID 320 332  
 FT CARBOHYD 283 283  
 FT CARBOHYD 296 296  
 FT CONFLICT 69 70  
 FT CONFLICT 147 148  
 FT SEQUENCE 448 AA; 50180 MW; 19FCAS1FDA328003 CRC64;

Query Match 10.4%; Score 209.5; DB 1; Length 448;  
 Best Local Similarity 28.8%; Pred. No. 1e-08;  
 Matches 65; Conservative 20; Mismatches 72; Indels 69; Gaps 14;

QY 138 LACGGGSG-----RPGSGNGHSGDG-----SROGDSGCRHMGY---QGPPLCTD- 179  
 DB 111 LICEFGYQPMDESNOCVVDVDECATSHOCNPTQICINTEGGYTCSTGQYMLLEGO-CIDI 169  
 QY 180 --CMDDGFSSLRNETHSICITACDESCKTCGSLTNRDGCECEVGVWLDG--ACVDVDECA 235  
 DB 170 DECRYGT-----CQQLCANVPG--SYSC-TCPGFTLNLDGNSCDVNECA 212  
 QY 236 AEPPPCSAOFCKNANGSYTCECDSSCGVGTGPGNCKEISGYAREHG--OCADYDE 293  
 DB 213 TENP---CVQTCVNTYSFIC-----RCDPGLELEEDGVCNDSMD 250  
 QY 294 CSLAEKTCVKRNENCYTPGYSVVCVCPDGEFETED--ACVPAARE 337  
 DB 251 CSFSEFLC---QHCCVNOPGTYPCSCPXYILLDNNSCODINECE 293

RESULT 43  
 VS41\_GIALA  
 ID VS41\_GIALA STANDARD; PRT; 687 AA.  
 AC P92127;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Variant-specific surface protein VSP4A1 precursor (CRISP-90).  
 OS Giardia lamblia (Giardia intestinalis).  
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.

OX NCBI\_TaxID=5741;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=O2-4A1.  
 RX MEDLINE=97321554; PubMed=9178264;  
 RA Papanastasiou P., Bruderer T., Li Y., Bommei C., Koehler P.;  
 RT "Primary structure and biochemical properties of a variant-specific  
 surface protein of Giardia.";  
 RL Mol. Biochem. Parasitol. 86:13-27(1997).  
 RP CHARACTERIZATION.  
 RX MEDLINE=97233006; PubMed=9078242;  
 RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;  
 RT "The variant-specific surface protein of Giardia, VSP4A1, is a  
 glycosylated and palmitoylated protein.";  
 RL Biochem. J. 322:49-56(1997).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE  
 PLASMA MEMBRANE.  
 CC -1- O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC  
 AT THE REDUCING TERMINUS.  
 CC -1- PTM: PALMITOYLATED.  
 CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: Z83743; CAB06038.1; -  
 DR GlycoSuiteDB: P92127; -  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR002174; Furin-like.  
 DR SMART: SM00181; EGF\_2.  
 DR SMART: SM00001; EGF\_Like; 1.  
 DR SMART: SM00261; FU; 3.  
 KW Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;  
 KW Signal.  
 FT SIGNAL 1 14  
 FT CHAIN 15 687  
 FT DOMAIN 15 660  
 FT TRANSMEM 661 681  
 FT DOMAIN 682 687  
 FT SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Query Match 10.4%; Score 209.5; DB 1; Length 687;  
 Best Local Similarity 23.6%; Pred. No. 1.5e-08;  
 Matches 70; Conservative 22; Mismatches 100; Indels 105; Gaps 14;

QY 129 SPGTYGPDCLACGGGSGR-PCSGNGHSGDGSROGDSGCRHMGYGP-LCTDCMDGYFS 186  
 DB 133 TPSEGSSECIICMPDITDRNGVAVANCA-----TCTAPASSTGATCTECMAGTYK 183  
 QY 187 SLRNETHSICITACDESCKTCGSLTNRDGCECEVGVWLDGACVVDVDECAAEPP----- 240  
 DB 184 KSDTE---CAACHSDCATCSGEANNOCTSCETGYLKSNOCEVNTNTHYPPDTSMT 239  
 QY 241 -----CSAOF-----CKNAN-----GSYCEECSSCGVGTG----- 268  
 DB 240 CVACTVLDANCAATCSFDSATAKGKCLTNSNKIRPTITLDGTSTCYE--NSTAGCGAGANE 297  
 QY 269 -----EGPGNCKECISGYAREHGQCAD-- 290  
 DB 298 LFMKEDGSACLLCGDTKEASNDKGVANCRCTKKNANDSPPTCTACLDGYFLERGSCITTYC 357  
 QY 291 VDEC-SLAETKTCVKRNENC-----YNTPGSYVCV-CPPGFEETEDACVPAAREATE 340  
 DB 358 ADNCATCSSEATTEEDCKICKAGFFLASPGEGKCSISDNTNGGIDGC-----ABCTK 409

RESULT	44		
CRB_DROME	CRB_DROME	STANDARD:	PRT: 2139 AA.
ID	CRB_DROME		
AC	P10040;		
DT	01-MAR-1989 (Rel. 10, Created)		
DT	01-MAY-1991 (Rel. 18, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Crumbs protein precursor (95F).		
GN	CRB.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OX	Ephydroidea; Drosophilidae; Drosophila.		
RN	NCBI_TaxID=7227;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=OREGON-R; TISSUE=Embryo;		
RA	MEDLINE=90263104; PubMed=2344615;		
RA	Tepass U., Theres C., Knust E.;		
RT	"Crumbs encodes an EGF-like protein expressed on apical membranes of		
RT	Drosophila epithelial cells and required for organization of		
RL	epithelia.";		
RL	Cell 61:787-799(1990).		
RN	[2]		
RP	SEQUENCE OF 1663-1955 FROM N.A.		
RP	TISSUE=Embryo;		
RX	MEDLINE=87218537; PubMed=3107986;		
RA	Knust E., Dietrich U., Tepass U., Bremer K.A., Weigelt D.,		
RA	Vassalli H., Campos-Ortega J.A.;		
RT	"EGF homologous sequences encoded in the genome of Drosophila		
RT	melanogaster, and their relation to neurogenic genes.";		
EMBO	J. 6:761-766(1987).		
CC	-I- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,		
CC	POLARITY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL		
CC	POLARITY. IT MAY ACT AS A SIGNAL.		
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-I- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.		
CC	-I- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL: M33753; AAA28428.1; ALT_SEQ.		
DR	EMBL: X05144; CA28793.1; .		
DR	PIR: B26637; B26637.		
DR	PIR: A35672; A35672.		
DR	HSSP: P00740; IIXA.		
DR	FlyBase: FBgn0000368; crb.		
DR	InterPro: IPR000152; Asx_hydroxyl.		
DR	InterPro: IPR000561; EGF-like.		
DR	InterPro: IPR000742; EGF_2		
DR	InterPro: IPR001881; EGF-Ca.		
DR	InterPro: IPR001388; EGF-II.		
DR	InterPro: IPR001791; laminin_G.		
DR	Pfam: PF00008; EGF_26.		
DR	Pfam: PF00054; laminin_G_3.		
DR	PRINTS: PR00010; EGFLDOD.		
DR	SMART: SM00179; EGF_CA_11.		
DR	SMART: SM00001; EGF_Like_16.		
DR	SMART: SM00282; Lamg; 3.		
DR	PROSITE: PS00010; ASX_HYDROXYL; 15.		
DR	PROSITE: PS00022; EGF_1; 26.		
DR	PROSITE: PS01186; EGF_2; 17.		
DR	PROSITE: PS01187; EGF_CA_15.		
DR	PROSITE: PS50025; LAM_G_DOMAIN; 3.		
DR	Differentiation; Repeat; EGF-like domain; Transmembrane;		
DR	Glycoprotein; Signal; Phosphorylation.		
FW			

---

FT	SIGNAL	1	90	
FT	CHAIN	91	2139	
FT	DOMAIN	91	2084	
FT	TRANSMEM	2085	2111	
FT	TRANSMEM	2112	2139	
FT	DOMAIN	267	303	
FT	DOMAIN	306	343	
FT	DOMAIN	348	386	
FT	DOMAIN	388	425	
FT	DOMAIN	427	463	
FT	DOMAIN	464	500	
FT	DOMAIN	501	532	
FT	DOMAIN	545	581	
FT	DOMAIN	582	611	
FT	DOMAIN	609	646	
FT	DOMAIN	648	685	
FT	DOMAIN	687	723	
FT	DOMAIN	725	761	
FT	DOMAIN	763	800	
FT	DOMAIN	802	838	
FT	DOMAIN	840	902	
FT	DOMAIN	904	940	
FT	DOMAIN	942	978	
FT	DOMAIN	980	1021	
FT	DOMAIN	1023	1205	
FT	DOMAIN	1207	1243	
FT	DOMAIN	1250	1480	
FT	DOMAIN	1481	1517	
FT	DOMAIN	1558	1758	
FT	DOMAIN	1759	1795	
FT	DOMAIN	1797	1833	
FT	DOMAIN	1835	1871	
FT	DOMAIN	1874	1915	
FT	DOMAIN	1915	1951	
FT	DOMAIN	1953	1989	
FT	DOMAIN	1991	2029	
FT	DOMAIN	2030	2070	
FT	DISULFID	271	282	
FT	DISULFID	276	291	
FT	DISULFID	293	302	
FT	DISULFID	310	321	
FT	DISULFID	315	331	
FT	DISULFID	333	342	
FT	DISULFID	352	363	
FT	DISULFID	357	374	
FT	DISULFID	376	385	
FT	DISULFID	392	403	
FT	DISULFID	397	412	
FT	DISULFID	414	424	
FT	DISULFID	431	442	
FT	DISULFID	436	442	
FT	DISULFID	453	461	
FT	DISULFID	458	462	
FT	DISULFID	468	479	
FT	DISULFID	473	488	
FT	DISULFID	490	499	
FT	DISULFID	505	515	
FT	DISULFID	509	520	
FT	DISULFID	522	531	
FT	DISULFID	549	562	
FT	DISULFID	556	569	
FT	DISULFID	571	580	
FT	DISULFID	586	597	
FT	DISULFID	591	602	
FT	DISULFID	604	610	
FT	DISULFID	618	624	
FT	DISULFID	613	634	
FT	DISULFID	618	634	
FT	DISULFID	636	645	
FT	DISULFID	652	664	
FT	DISULFID	659	673	
FT	DISULFID	675	684	
FT	DISULFID	691	702	
FT	DISULFID	716	711	
FT	DISULFID	713	722	

CRUMBS PROTEIN,	
EXTRACELLULAR (POTENTIAL).	
POTENTIAL.	
CYTOPLASMIC (POTENTIAL).	
EGF-LIKE 1.	
EGF-LIKE 2.	
EGF-LIKE 3.	
EGF-LIKE 4.	
EGF-LIKE 5.	
EGF-LIKE 6.	
EGF-LIKE 7.	
EGF-LIKE 8.	
EGF-LIKE 9.	
EGF-LIKE 10.	
EGF-LIKE 11.	
EGF-LIKE 12.	
EGF-LIKE 13.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 14.	
EGF-LIKE 15.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 16.	
EGF-LIKE 17.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 18.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 19.	
LAMININ G-LIKE 1.	
EGF-LIKE 20.	
LAMININ G-LIKE 2.	
EGF-LIKE 21.	
LAMININ G-LIKE 3.	
EGF-LIKE 22.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 23.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 24.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 25.	
EGF-LIKE 26.	
EGF-LIKE 27.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 28.	
CALCIUM-BINDING (POTENTIAL).	
EGF-LIKE 29.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	
BY SIMILARITY.	



```

FT DISULFID 729 740 BY SIMILARITY.
FT DISULFID 734 749 BY SIMILARITY.
FT DISULFID 751 760 BY SIMILARITY.
FT DISULFID 767 778 BY SIMILARITY.
FT DISULFID 772 787 BY SIMILARITY.
FT DISULFID 789 799 BY SIMILARITY.
FT DISULFID 806 817 BY SIMILARITY.
FT DISULFID 811 826 BY SIMILARITY.
FT DISULFID 828 837 BY SIMILARITY.
FT DISULFID 844 855 BY SIMILARITY.
FT DISULFID 849 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 908 919 BY SIMILARITY.
FT DISULFID 913 928 BY SIMILARITY.
FT DISULFID 930 939 BY SIMILARITY.
FT DISULFID 946 957 BY SIMILARITY.
FT DISULFID 952 966 BY SIMILARITY.
FT DISULFID 968 977 BY SIMILARITY.
FT DISULFID 984 995 BY SIMILARITY.
FT DISULFID 989 1009 BY SIMILARITY.
FT DISULFID 1011 1020 BY SIMILARITY.
FT DISULFID 1211 1222 BY SIMILARITY.
FT DISULFID 1216 1231 BY SIMILARITY.
FT DISULFID 1233 1242 BY SIMILARITY.
FT DISULFID 1485 1496 BY SIMILARITY.
FT DISULFID 1490 1505 BY SIMILARITY.
FT DISULFID 1507 1516 BY SIMILARITY.
FT DISULFID 1763 1774 BY SIMILARITY.
FT DISULFID 1768 1783 BY SIMILARITY.
FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.
FT DISULFID 1979 1988 BY SIMILARITY.
FT DISULFID 1995 2008 BY SIMILARITY.
FT DISULFID 2002 2017 BY SIMILARITY.
FT DISULFID 2019 2028 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match Best Local Similarity 10.48; Score 209; DB 1; Length 2139;  
Matches 63; Conservative 20; Mismatches 78; Indels 78; Gaps 14;

```

QY 125 KVCSPGTYGPGCLACGSGQR-----CSG-NGHSCGSGSGSGSGS-CRCHMGYQGP 175
DB 519 KCVCKPSSSGTNCOTSTGDCASALALTPINCNAINGKLNKGTCSMNGTHTCAVGYSGD 578
QY 176 LC-----TDCMDGYFSSLRNE---THSICGTAC-----DESCCTC 206
DB 579 RCEKAENCSPLNCOEP-WVCVQNCCLCPENKYVNCOCATOPONGGECVDPNGDYECCT 637
QY 207 SGLNRGCGEVEGVWLDGACAPRPPCSAOPFKMANGSTTCECCDSSCVGC 266
DB 638 RGMWGRTCGN-----DVDECTLHPKICGNG-ICKNEKGSYKC----- 673

```

```

QY 267 TGEPRGNKECISGAREHGOC-ADVDECSLAETKVCARNKNCYTPGSYVCVCPDGE 324
DB 674 -----YCTPPTGTGVH--CDSVDVDC--LSFPCCL-NGATCKHKNATYACVCPGTE 718

RESULT 45
ID FBIL_CRIGR STANDARD: PRT: 443 AA.
AC 05058;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE EGF-containing fibulin-like extracellular matrix protein 2 precursor
DE (fibulin-4) (FBLN-4) (H411 protein).
CN EFEMP2 OR FBILN4.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Heine H., Delude R.L., Monks B., Golenbock D.T.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF046870; AAC03101.1; -.
DR HSPSP: P00736; IAPQ.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00008; EGF_4.
DR SMART: SM00179; EGF_CA_4.
DR SMART: SM00001; EGF_Like_2.
DR PROSITE: PS00010; ASX_HYDROXYL_4.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 4.
DR PROSITE: PS01187; EGF_CA_6.
KW Repeat: EGF-like domain; Calcium-binding; Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 443
FT FT 1 443
FT FT 26 443
FT FT 123 163
FT FT 164 202
FT FT 203 242
FT FT 243 282
FT FT 283 328
FT FT 127 140
FT FT DISULFID 134 149
FT FT DISULFID 151 162
FT FT DISULFID 168 177
FT FT DISULFID 173 186
FT FT DISULFID 188 201
FT FT DISULFID 207 217
FT FT DISULFID 213 226
FT FT DISULFID 228 241
FT FT DISULFID 247 258
FT FT DISULFID 254 267
FT FT DISULFID 269 281
FT FT DISULFID 287 300
FT FT DISULFID 294 309
FT FT DISULFID 315 327
FT FT DISULFID 315 327

```

FT CAROHD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CAROHD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 443 AA; 494322 MW; 0BCFED5D7323D9E5F CRC64;

Query Match  
 Best Local Similarity 37.0%; Score 208; DB 1; Length 443;  
 Matches 47; Conservative 12; Mismatches 56; Indels 12; Gaps 5;

OY 216 ECEVGG--WVLDGACVYDVECAEPPPCSAQPCNKANGSYTCEDSSCVG-CTGEGP- 271  
 DB 38 ECTDGTGEMDADSGHCRDVEDCLTTPACKGEMKINHGYGLCLPSAAVINLHGEGR 97  
 OY 272 -----GNCKECISGY-AREHGQCADVDECSIAEKTCKRKNENYTPGSSYCVCPDGE 324  
 DB 98 PVPRAQHNPCCPPGYEPDEQESCVDECAQALHDC-RPSQCHNLPGSYQCTCPDGYR 156  
 OY 325 ETEDACV 331  
 DB 157 KVGPECV 163

RESULT 46  
 FB4\_MOUSE  
 ID FB4\_MOUSE STANDARD; PRT; 443 AA.  
 AC Q9WVJ9;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE EGF-containing fibulin-like extracellular matrix protein 2 precursor  
 DE (Fibulin-4) (FBLN-4) (Mutant p53 binding protein 1).  
 GN EFEMP2 OR FBLN4 OR MBP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=9308589; PubMed=10380882;  
 RA Gallagher W.M., Argentini M., Sierra V., Bracco L., Debussche L.,  
 RA Conseiller E.;  
 RT "MBP1: a novel mutant p53-specific protein partner with oncogenic  
 properties.";  
 RL Oncogene 18:3608-3616(1999).  
 CC -I- SUBUNIT: BINDS PREFERENTIALLY TO P53 MUTANTS.  
 CC -I- SUBCELLULAR LOCATION: Secreted.  
 CC -I- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF104223; AAD45219.1; .  
 DR HSSP; P00736; IAPQ.  
 DR MGD; MGI:1891209; Efemp2.  
 DR InterPro; IPR000152; ASX\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF-Ca.  
 DR InterPro; JPR001491; Thrombomodulin.  
 DR Pfam; PFO0008; EGF; 4.  
 DR PRINTS; PR00907; THROMBODULN.  
 DR SMART; SM00179; EGF\_CA; 4.  
 DR SMART; SM00001; EGF\_Like; 2.  
 DR PROSITE; PS00020; ASX\_HYDROXYL; 4.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS01187; EGF\_CA; 6  
 KW Repeat; EGF-like domain; Calcium-binding; Glycoprotein; Signal.

FT SIGNAL 1 25  
 FT CHAIN 26 443  
 FT FT  
 FT DOMAIN 36 81  
 FT DOMAIN 123 163  
 FT DOMAIN 164 202  
 FT DOMAIN 203 242  
 FT DOMAIN 243 282  
 FT DOMAIN 283 328  
 FT DOMAIN 328 358  
 FT DISULFID 127 140  
 FT DISULFID 134 149  
 FT DISULFID 151 162  
 FT DISULFID 168 177  
 FT DISULFID 173 186  
 FT DISULFID 188 201  
 FT DISULFID 207 217  
 FT DISULFID 213 226  
 FT DISULFID 228 241  
 FT DISULFID 247 258  
 FT DISULFID 254 267  
 FT DISULFID 269 281  
 FT DISULFID 287 300  
 FT DISULFID 294 309  
 FT DISULFID 315 327  
 FT CAROHD 198 198  
 FT CAROHD 394 394  
 SQ SEQUENCE 443 AA; 49425 MW; 4969C0328A23BD88 CRC64;

Query Match  
 Best Local Similarity 10.3%; Score 207; DB 1; Length 443;  
 Matches 47; Conservative 11; Mismatches 57; Indels 12; Gaps 5;

OY 216 ECEVGG--WVLDGACVYDVECAEPPPCSAQPCNKANGSYTCEDSSCVG-CTGEGP- 271  
 DB 38 ECTDGTGEMDADSGHCRDVEDCLTTPACKGEMKINHGYGLCLPSAAVINLHGEGR 97  
 OY 272 -----GNCKECISGY-AREHGQCADVDECSIAEKTCKRKNENYTPGSSYCVCPDGE 324  
 DB 98 PPAHAQNPCCPPGYEPDEQESCVDECAQALHDC-RPSQCHNLPGSYQCTCPDGYR 156  
 OY 325 ETEDACV 331  
 DB 157 KVGPECV 163

RESULT 47  
 FP2\_MYTGA  
 ID FP2\_MYTGA STANDARD; PRT; 473 AA.  
 AC Q25464;  
 DT 30-MAY-2000 (Rel. 39; Created)  
 DT 30-MAY-2000 (Rel. 39; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE Adhesive plaque matrix protein 2 precursor (Foot protein 2) (MGFP2)  
 DE (MGFP-2).  
 GN FP2.  
 OS Mytilus galloprovincialis (Mediterranean mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=29158;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Foot;  
 RC TISSUE=Foot;  
 RX MEDLINE=95204464; PubMed=7896812;  
 RX Inoue K., Takeuchi Y., Miki D., Odo S.;  
 RA "Mussel adhesive plaque protein gene is a novel member of epidermal  
 RT growth factor-like gene family.";  
 RL J. Biol. Chem. 270:6698-6701(1995).  
 CC -I- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
 CC -I- SUBCELLULAR LOCATION: Secreted.



CC	modified and this statement is not removed.
CC	entitles requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; U48246; AAC72252.1; .
DR	HSSP; P07204; IADX.
DR	InterPro; IPR0000152; Asx_hydroxyl.
DR	InterPro; IPR0000561; EGF-like.
DR	InterPro; IPR001881; EGF_Ca.
DR	InterPro; IPR001791; Laminin_G.
DR	InterPro; IPR001007; TSPN.
DR	Pfam; PF002210; TSPN; 1.
DR	Pfam; PF00093; vwc; 3.
DR	SMART; SM00179; EGF_CA; 2.
DR	SMART; SM00001; EGF_Like; 4.
DR	SMART; SM00282; Lamg; 1.
DR	SMART; SM00210; TSPN; 1.
DR	SMART; SM00214; VWC; 4.
DR	PROSITE; PS00010; ASX_HYDROXYL; 3.
DR	PROSITE; PS00022; EGF_1; 1.
DR	PROSITE; PS01186; EGF_2; 3.
DR	PROSITE; PS01187; EGF_CA; 3.
DR	PROSITE; PS01208; VWFC; 2.
KW	Glycoprotein; EGF-Like domain; Repeat; signal.
FT	SIGNAL 1 16
FT	CHAIN 17 810
FT	DOMAIN 81 230
FT	DOMAIN 273 331
FT	DOMAIN 335 390
FT	DOMAIN 391 433
FT	DOMAIN 434 475
FT	DOMAIN 476 516
FT	DOMAIN 515 547
FT	DOMAIN 549 595
FT	DOMAIN 596 631
FT	DOMAIN 632 687
FT	DOMAIN 692 750
FT	DOMAIN 752 807
FT	DISULFID 395 407
FT	DISULFID 401 416
FT	DISULFID 418 432
FT	DISULFID 438 451
FT	DISULFID 445 460
FT	DISULFID 462 474
FT	DISULFID 480 493
FT	DISULFID 487 502
FT	DISULFID 504 515
FT	DISULFID 519 529
FT	DISULFID 523 535
FT	DISULFID 537 546
FT	DISULFID 553 566
FT	DISULFID 560 575
FT	DISULFID 577 594
FT	DISULFID 600 613
FT	DISULFID 607 622
FT	DISULFID 624 630
FT	CARBOHYD 40 40
FT	CARBOHYD 53 53
FT	CARBOHYD 83 83
FT	CARBOHYD 224 224
FT	CARBOHYD 294 294
FT	CARBOHYD 372 372
FT	CARBOHYD 511 511
FT	CARBOHYD 562 562
FT	CARBOHYD 609 609
FT	CARBOHYD 708 708
SO	SEQUENCE 810 AA; 89212 MW; 46F09C466AF9AB0B CRC64;

```

Matches 65; Conservative 17; Mismatches 72; Indels 67; Gaps 13;
Qy 126 VCCS-PGTYRGDCLACGGSOR-----PCSGHGCSGDSRGDSRCRHMRYOGELCTDC 180
Db 450 VCVNLPGYLRCDYCV--GYLRVDFSCSTEHDDC-GSGGHNCNDKNAITNTVQGISCT-C 504
Qy 181 MDGFSSLRNETHSICTA-CDESC--TC-----SGLTNRDGCCEVGNWLDGA 227
Db 505 QPGYGVN-----GTICKAFCEGCRYGGTCAVAPNKVCVPSGPTGSHCK----- 548
Qy 228 CVDVDECAEPPPCSAOFCCKNANGSYTCECDSSCVCTGEGPQNCECTISG-----Y 281
Db 549 --DIDCEAGEVEGHCHNYSRCVNLPEMYHC-----ECRSGFHDGTY 587
Qy 282 AREHGOCADVEDCSLAERTCVKRNKNCYNTPGSYVCYCPDG 322
Db 508 SLSGESCIDIDECALRTHTC-WNDSACINLAGFPCLCPSG 627

RESULT 49
NTC4_MOUSE
ID NTC4_MOUSE STANDARD; PRT; 1964 AA.
AC P31695; 062389;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch homolog protein 4 precursor (Transforming
DE protein INT-3).
GN NOTCH4 OR INT3 OR INT-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid:10090;
RN [1]
RZ SEQUENCE FROM N.A.
RX MEDLINE-92194507; PubMed-1312643;
RA Robbins J., Blondel B.J., Gallahan D., Gallahan R.;
RT "Mouse mammary tumor gene Int-3: a member of the notch gene family
RT transforms mammary epithelial cells.";
RL J. Virol. 66:2594-2599(1992).
RN [2]
RZ REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE-97294599; PubMed-9150355;
RA Gallahan D., Gallahan R.;
RT "The mouse mammary tumor associated gene INT3 is a unique member of
RT the NOTCH gene family (NOTCH4).";
RL Oncogene 14:1883-1890(1997).
RN [3]
RZ SEQUENCE FROM N.A.
RX TISSUE=Lung, and Testis;
RA MEDLINE-96281668; PubMed-6681805;
RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT "Notch4/Int-3, a mammary proto-oncogene, is an endothelial
RT cell-specific mammalian Notch gene.";
RL Development 122:2251-2259(1996).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 5 AMR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M80456; AAB38377.1; -
DR EMBL: U43691; AAC52630.1; -
DR PIR: A38072; TNAVY3.
DR HSSP: P08709; 1BF9.

```

DR MGD; MGI:107471; Notch4.  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR000742; EGF\_2.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001438; EGF\_11.  
DR InterPro: IPR000800; Notch.  
DR Pfam: PF000023; ank; 6.  
DR Pfam: PF00008; EGF; 27.  
DR Pfam: PF00066; notch; 2.  
DR PRINTS; PR01415; ANKYRIN.  
DR PRINTS; PR0010; EGFBL0D.  
DR PRINTS; PR01452; NOTCH.  
DR SMART; SM00248; ANK; 5.  
DR SMART; SM00179; EGF\_Ca; 11.  
DR SMART; SM00001; EGF\_Like; 15.  
DR SMART; SM00004; NL; 2.  
DR PROSITE; PS50088; ANK\_REPEAT; 5.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 11.  
DR PROSITE; PS00022; EGF\_1; 28.  
DR PROSITE; PS01186; EGF\_2; 21.  
DR PROSITE; PS01187; EGF\_Ca; 9.  
KW Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;  
KW Glycoprotein; Proto-oncogene; ANK repeat; Signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 1964 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.  
FT DOMAIN 21 1443 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1444 1464 POTENTIAL.  
FT DOMAIN 1465 1964 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 21 60 EGF-LIKE 1.  
FT DOMAIN 61 112 EGF-LIKE 2.  
FT DOMAIN 115 152 EGF-LIKE 3.  
FT DOMAIN 153 189 EGF-LIKE 4.  
FT DOMAIN 191 229 EGF-LIKE 5.  
FT DOMAIN 231 271 EGF-LIKE 6.  
FT DOMAIN 273 309 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 311 350 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 352 388 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 389 427 EGF-LIKE 10.  
FT DOMAIN 429 470 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 472 508 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 510 546 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 548 584 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 586 622 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 622 656 EGF-LIKE 16.  
FT DOMAIN 658 686 EGF-LIKE 17.  
FT DOMAIN 688 724 EGF-LIKE 18.  
FT DOMAIN 726 762 EGF-LIKE 19.  
FT DOMAIN 764 800 EGF-LIKE 20.  
FT DOMAIN 803 839 EGF-LIKE 21.  
FT DOMAIN 841 877 EGF-LIKE 22.  
FT DOMAIN 878 924 EGF-LIKE 23.  
FT DOMAIN 926 962 EGF-LIKE 24.  
FT DOMAIN 964 1000 EGF-LIKE 25.  
FT DOMAIN 1002 1040 EGF-LIKE 26.  
FT DOMAIN 1042 1081 EGF-LIKE 27.  
FT DOMAIN 1083 1122 EGF-LIKE 28.  
FT DOMAIN 1126 1167 EGF-LIKE 29.  
FT REPEAT 1168 1208 LIN/NOTCH 1.  
FT REPEAT 1209 1242 LIN/NOTCH 2.  
FT REPEAT 1243 1282 LIN/NOTCH 3.  
FT REPEAT 1282 1318 ANK 1.  
FT REPEAT 1318 1354 ANK 2.  
FT REPEAT 1354 1390 ANK 3.  
FT REPEAT 1390 1426 ANK 4.  
FT REPEAT 1426 1462 ANK 5.  
FT DISULFID 25 38 BY SIMILARITY.  
FT DISULFID 32 48 BY SIMILARITY.  
FT DISULFID 30 59 BY SIMILARITY.  
FT DISULFID 77 1028 BY SIMILARITY.  
FT DISULFID 71 100 BY SIMILARITY.  
FT DISULFID 102 111 BY SIMILARITY.  
FT DISULFID 119 130 BY SIMILARITY.  
FT DISULFID 124 140 BY SIMILARITY.  
FT DISULFID 142 151 BY SIMILARITY.  
FT DISULFID 157 168 BY SIMILARITY.  
FT DISULFID 162 177 BY SIMILARITY.  
FT DISULFID 179 188 BY SIMILARITY.  
FT DISULFID 195 208 BY SIMILARITY.  
FT DISULFID 202 217 BY SIMILARITY.  
FT DISULFID 219 228 BY SIMILARITY.  
FT DISULFID 235 246 BY SIMILARITY.  
FT DISULFID 240 259 BY SIMILARITY.  
FT DISULFID 261 270 BY SIMILARITY.  
FT DISULFID 277 288 BY SIMILARITY.  
FT DISULFID 282 297 BY SIMILARITY.  
FT DISULFID 299 308 BY SIMILARITY.  
FT DISULFID 315 329 BY SIMILARITY.  
FT DISULFID 323 338 BY SIMILARITY.  
FT DISULFID 340 349 BY SIMILARITY.  
FT DISULFID 356 367 BY SIMILARITY.  
FT DISULFID 361 376 BY SIMILARITY.  
FT DISULFID 378 387 BY SIMILARITY.  
FT DISULFID 393 404 BY SIMILARITY.  
FT DISULFID 398 415 BY SIMILARITY.  
FT DISULFID 417 426 BY SIMILARITY.  
FT DISULFID 433 449 BY SIMILARITY.  
FT DISULFID 443 458 BY SIMILARITY.  
FT DISULFID 460 469 BY SIMILARITY.  
FT DISULFID 476 487 BY SIMILARITY.  
FT DISULFID 481 496 BY SIMILARITY.  
FT DISULFID 498 507 BY SIMILARITY.  
FT DISULFID 514 525 BY SIMILARITY.  
FT DISULFID 519 545 BY SIMILARITY.  
FT DISULFID 536 563 BY SIMILARITY.  
FT DISULFID 552 572 BY SIMILARITY.  
FT DISULFID 574 583 BY SIMILARITY.  
FT DISULFID 590 601 BY SIMILARITY.  
FT DISULFID 595 610 BY SIMILARITY.  
FT DISULFID 612 621 BY SIMILARITY.  
FT DISULFID 626 637 BY SIMILARITY.  
FT DISULFID 631 646 BY SIMILARITY.  
FT DISULFID 648 655 BY SIMILARITY.  
FT DISULFID 662 669 BY SIMILARITY.  
FT DISULFID 664 674 BY SIMILARITY.  
FT DISULFID 676 685 BY SIMILARITY.  
FT DISULFID 692 703 BY SIMILARITY.  
FT DISULFID 697 712 BY SIMILARITY.  
FT DISULFID 714 723 BY SIMILARITY.  
FT DISULFID 730 741 BY SIMILARITY.  
FT DISULFID 735 750 BY SIMILARITY.  
FT DISULFID 752 761 BY SIMILARITY.  
FT DISULFID 768 779 BY SIMILARITY.  
FT DISULFID 773 788 BY SIMILARITY.  
FT DISULFID 799 807 BY SIMILARITY.  
FT DISULFID 807 818 BY SIMILARITY.  
FT DISULFID 812 827 BY SIMILARITY.  
FT DISULFID 829 838 BY SIMILARITY.  
FT DISULFID 845 856 BY SIMILARITY.  
FT DISULFID 850 865 BY SIMILARITY.  
FT DISULFID 867 876 BY SIMILARITY.  
FT DISULFID 882 893 BY SIMILARITY.  
FT DISULFID 897 912 BY SIMILARITY.  
FT DISULFID 914 923 BY SIMILARITY.  
FT DISULFID 930 941 BY SIMILARITY.  
FT DISULFID 935 950 BY SIMILARITY.  
FT DISULFID 952 961 BY SIMILARITY.  
FT DISULFID 968 979 BY SIMILARITY.  
FT DISULFID 973 988 BY SIMILARITY.  
FT DISULFID 990 999 BY SIMILARITY.  
FT DISULFID 1006 1019 BY SIMILARITY.  
FT DISULFID 1011 1028 BY SIMILARITY.  
FT DISULFID 1030 1039 BY SIMILARITY.

FT DISULFID 1046 1057 BY SIMILARITY.  
 FT DISULFID 1051 1069 BY SIMILARITY.

Query Match  
 Best Local Similarity 22.9%; Score 204.5; DB 1; Length 1964;  
 Matches 79; Conservative 28; Mismatches 95; Indels 143; Gaps 18;

OY 128 CSPGTVPDCLACQGGSSORPCSGNGHC-----154  
 DB 142 CEGMTGECQCLREDFCSANPCANGVCLATYPOIQCRPPGFGHNTCERDINECFLEPGP 201  
 OY 155 --SGDSRQDGS--CRCHMGVGPCT-----DCMDGFFSSLRNTHSICACDE 201  
 DB 202 CPQGTSCHTLGSYQCLCPVGGEGPQCKLRKACAPGSCSLNGTQCLVPGHSTFHLG-- 259  
 OY 202 SCKTC-SGLTNRDC-----GCEVGVV-----LDEGACV-----DVDECA 235  
 DB 260 ---LCPPGFTGLDCENMPDCAVHQCQNGATCLDGLDTYTCCLPKTWKMGDCSEDIDECE 316  
 OY 236 AE-PPPCSAOQFCKNANGSYTC-----ERC-----DSSVCCTGE-- 269  
 DB 317 ARGPFCRNGTQQTQNGAFHFCVCGSWGAGCEENLMDCAATCAFGSTCIDRVGSFSC 376  
 OY 270 --GPGN---C---KECIGYAREHGCA-----DVDECSIA 297  
 DB 377 LCPGRTGLCHLEDMCLSPCHVNAQCSNPLTGLTICOPYSGSTCHODLECOMA 436  
 OY 298 EK--TCVRKNENYNTPGSVVVCVCPFEETEDACVPAEAATE 340  
 DB 437 QQGPSPCEHGGSCINTPGSFNCLCLPGY--TGSRC-----EADHNE 475

RESULT 50  
 FBIL HUMAN STANDARD; PRT; 443 AA.

AC 095967: 075967;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE EGF-containing fibulin-like extracellular matrix protein 2 precursor (fibulin-4) (FBL-4) (UBH1 protein).  
 GN EFEMP2 OR FBLN4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=20068041; PubMed=10601734;  
 RA Gilly R., Timpl R., Kostka G.;  
 RT "Sequence, recombinant expression and tissue localization of two novel extracellular matrix proteins, fibulin-3 and fibulin-4.";  
 RL Matrix Biol. 18:469-480(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Zemel R., Shaul Y.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20435063; PubMed=10982184;  
 RA Katsanis N., Venable S., Smith J.R., Lupski J.R.;  
 RT "Isolation of a paralog of the Doyme honeycomb retinal dystrophy gene Hum. Genet. 106:66-72(2000).  
 RL Hum. Genet. 106:66-72(2000).  
 CC -1- SURCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AJ132819; CA010791.2; -  
 DR EMBL: AF093119; AAC62108.1; -  
 DR EMBL: AF109121; AAF65188.1; -  
 DR HSSP: P35355; 1EMN.  
 DR MIM: 604633; -  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF Ca.  
 DR InterPro: IPR001491; Thrbomodulin.  
 DR Pfam: PF00008; EGF; 4.  
 DR PRINTS: PR00907; THRBOMODULN.  
 DR SMART: SM00179; EGF\_CA; 4.  
 DR SMART: SM00001; EGF\_Like; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 4.  
 DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE: PS01186; EGF\_2; 4.  
 DR PROSITE: PS01187; EGF\_CA; 6.  
 DR Repeat: EGF-like domain; Calcium-binding; Glycoprotein; signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 443  
 FT DOMAIN 36 81  
 FT DOMAIN 123 163  
 FT DOMAIN 164 202  
 FT DOMAIN 203 242  
 FT DOMAIN 243 282  
 FT DOMAIN 283 328  
 FT DISULFID 127 140  
 FT DISULFID 134 149  
 FT DISULFID 151 162  
 FT DISULFID 168 177  
 FT DISULFID 173 186  
 FT DISULFID 188 201  
 FT DISULFID 207 217  
 FT DISULFID 213 226  
 FT DISULFID 228 241  
 FT DISULFID 247 258  
 FT DISULFID 254 267  
 FT DISULFID 269 281  
 FT DISULFID 287 300  
 FT DISULFID 294 309  
 FT DISULFID 315 327  
 FT CARBOHYD 198 198  
 FT CARBOHYD 394 394  
 FT CONFLICT 5 5  
 FT CONFLICT 44 51  
 FT CONFLICT 103 111  
 FT CONFLICT 294 294  
 FT CONFLICT 354 356  
 FT CONFLICT 355 355  
 SQ SEQUENCE 443 AA; 49391 MW; 9E9AC2393780D3B8 CRC64;

Query Match  
 Best Local Similarity 10.1%; Score 203; DB 1; Length 443;  
 Matches 46; Conservative 13; Mismatches 56; Indels 12; Gaps 5;

OY 216 ECEVGG--WVLDEGACVDVDECAEPPPCSAOQFCKNANGSYTCECDSSCVG-CGEGEP- 271  
 DB 38 ECTDGEWMPDSDQHCRDVENCLTIPACKGEMKCIHHYGYLCLPRSAVINDLHGEPP 97  
 OY 272 -----GNCKECISGV-AEHGQCAVDVDCSLAETKCVKKNENYNTPGSVVCPDGP 324  
 DB 98 PVPYPAQHNPCCPGYEPDDSDSDVDECAQALHDC-RPSODCHNLPSGYCTCPDGYR 156  
 OY 325 ETEDACV 331  
 DB 157 KIGPECV 163

Wed Sep 11 09:09:34 2002

us-09-905-075-2.top50.rsp

Page 61

Search completed: September 10, 2002, 11:13:02  
Job time: 219 sec

---





[illegible]

Db 132 QWLCSDSLKCCPAGTFGSPCLPFGGTERPCGGYQCEBEGTRGSGHCDCAAGYGEA 191  
 QY 177 CTFCDMGYFSSLSNETHSICITACDESKCTCSGLTNRDGCRCFCEYVLDGACVDECA 236  
 Db 192 CGCGGAGTGEAEENASHLYCSACFGPCARCSGSEBSCNLCCKGMALHLKCTVDIECGT 251  
 QY 237 EPPPCSAOFCCKNANGSYTCCECDSSCVGCTGEGNCKEJCISGVAREHQCADVDECSL 296  
 Db 252 EGANGCAGDFCVNTEGSEYECRDCAKACLGCMGAPGRCKKCSFGYQVGSCKLDVDECE- 310  
 QY 297 AETCYCRKNKNCNTPGSYVCYCPDPFEETEDACP-PAEA-EEATEGE 342  
 Db 311 -TEVCGENKCKCENTEGYRCICAECKQMEGICVKEQIPESAGFSEMTED 362

## RESULT 2

hypothetical protein F09E8.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T20656  
 R:Pericy, C.  
 submitted to the EMBL Data Library, May 1996  
 A:Reference number: Z19307  
 A:Accession: T20656  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-356 <TIL>  
 A:Cross-references: EMBL:Z73896; PIDN:CAA98055.1; GSPDB:GN00022; CESP:F09E8.2  
 A:Experimental source: clone F09E8  
 C:Genetics:  
 A:Gene: CESP:F09E8.2  
 A:Map position: 4  
 A:Introns: 39/3; 67/2; 99/3; 134/1; 165/3; 208/1; 241/1; 287/1

Query Match 26.9%; Score 540; DB 2; Length 356;  
 Best local Similarity 36.1%; Pred. No. 1.5e-27;  
 Matches 108; Conservative 43; Mismatches 110; Indels 38; Gaps 6;

QY 11 LPFLLLLPAPPAKKPRPCHRCRGIVDKFNGMGVDTAKKNGGNTAMEEKLKYS 70  
 Db 7 LIAVLIGANSQKEVTITKNEKRCFNPVSTFDGLKTAHNFAGDGTAMEEKNLKYKT 66  
 QY 71 SEIRLEILEGLCESS-----DFECNMLEAEDEHLLEAMWLKSEYDLE 116  
 Db 67 SEIRLIEVLEGVCKKSSLPAMDPMGIALEIEFKCSQOLEKHEETIEFY--YNOGINNMS 124  
 QY 117 EMFCVTKLVKVCSPGTGPDCLACGGGSGRP--CSGNGHCSGDSRQGGSCRCRHMGYOG 174  
 Db 125 NMVLCEQLKICCPDGHFGKKECEQCPGISERKADVCFGKSGCHGSGRSGKCKCETGYTG 184  
 QY 175 PLCTDMDGFFSSLRNETHSICITACDESC-KTCSGLTNRCCGCEYGVLDGACVADVE 233  
 Db 185 NLCRVCDIEFFESKRVQGVCKKCHGCGICVCSSESSCKSKCKMGLTEGCGADVNE 244  
 QY 234 CAEPPEPCSAOFCCKNANGSYTCCECDSSCVGCTGEGNCKEJCISGVAREHQCADVD 292  
 Db 245 CQNESACTKHEICVMTVGSFKCE-----CKE---GYKKDDQDNQOPD 284

## RESULT 3

fibrillin 1 - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 11-Jan-2000  
 C:Accession: A55567  
 R:Ristola, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.  
 Genomics 23, 480-485, 1994  
 A:Title: Sequence of the coding region of the bovine fibrillin cDNA and localization to  
 A:Reference number: A55567; MUID:95137557  
 A:Accession: A55567  
 A:Status: preliminary

A:Molecule type: mRNA  
 A:Residues: 1-2871 <TIL>  
 A:Cross-references: GB:I28748; NID:9508427; PIDN:AAA74122.1; PID:9508428  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 F:1201-1236/Domain: EGF homology <EGF>

Query Match 13.5%; Score 270; DB 2; Length 2871;  
 Best local Similarity 24.8%; Pred. No. 1.3e-09;  
 Matches 107; Conservative 33; Mismatches 128; Indels 164; Gaps 25;

QY 19 PPAPPAKKPRPC-----HRCGLYDKF-----NQGAVDTAKNFGGNTA 59  
 Db 1141 PRQHLAPNISACIDINECELSAHLCPHRCVNLIGKYOCACNPYHSPTDRLF----- 1194  
 QY 60 WKEKTLSEKESSEIRLEILEGLCES-----SDEFCN-----QMLEAGEHLEAMWLQ 108  
 Db 1195 -----CVIDIDCSITMNGCETFCNTEGSEYCSGCPGFPALMPDOR----- 1234  
 QY 109 KSEYDLEFEMFCVTKLVKVC-----CS--PGTYGPDCLACOG-----GSORPCSGNGHC----- 154  
 Db 1235 --SCTDIDE--CEDNPNIDGGGCTNIPGEY--RCLCYGFMASEDMKTCVVDNEDLMP 1288  
 QY 155 ----SGD--GSROGDSRCRHMGYOGLPL-----CTDCMDGYFSSLRNETHSICITACDESKCT 205  
 Db 1289 NCLSGTCENTGSGSFICHDMDYSGKKGKTCGTDJNECEIGHNCDRAHVCTNTAGSFK- 1347  
 QY 206 GSGLTNRDCGECEYGVLDGACVADVECAEPPEPCSAOFCCKNANGSYTC----- 256  
 Db 1348 CS-----CSPGWIGIGIKCTDIDECNSGTHMCSQAHADCKNTMGSTRCLCKEYTGDD 1398  
 QY 257 -----EBC-----DSCVGCCTG-----EGP----- 271  
 Db 1399 GFTCTDLDCSENLNCGNGCCLNAPGGRCECDMGFVSPADKACEDIDECSLPNICVF 1458  
 QY 272 GNCK-----ECISGYA--REHGCAVDDECSLAETKCYKKNENGYTPRSGYVCVD 321  
 Db 1459 GTCHNLPGLFRCCECEGYELDRGGGCTDYNBC-IDPTTCI--SGNCVMTPGSYTCDP 1515  
 QY 322 GFE--ETEDACV 331  
 Db 1516 DFELNPTRYGCV 1527

## RESULT 4

fibrillin 1 precursor - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Jun-1995 #sequence\_revision 25-Apr-1997 #text\_change 21-Jul-2000  
 C:Accession: A47221; I54355; S17064; I59574; S17062; S6211L; A34198  
 R:Corson, G.M.; Chabberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y.  
 Genomics 17, 476-484, 1993  
 A:Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain stru  
 A:Reference number: A47221; MUID:94010947  
 A:Accession: A47221  
 A:Molecule type: mRNA  
 A:Residues: 1-337, 'T', 339-1029 <COR>  
 A:Cross-references: GB:X63556  
 R:Perleir, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Panglman, T.; Bo  
 Hum. Mol. Genet. 2, 961-968, 1993  
 A:Title: Genomic organization of the sequence coding for fibrillin, the defective gen  
 A:Reference number: I54355; MUID:93372860  
 A:Accession: I54355  
 A:Molecule type: mRNA  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Residues: 132-3002 <PER>  
 A:Cross-references: GB:LI9923; NID:9306745; PIDN:AA02036.1; PID:9306746  
 R:Maslen, C.L.; Corson, G.M.; Maddox, B.K.; Ghanville, R.W.; Sakai, L.Y.  
 Nature 352, 334-337, 1991  
 A:Title: Partial sequence of a candidate gene for the Marfan syndrome.  
 A:Reference number: S17064; MUID:91304568  
 A:Accession: S17064  
 A:Molecule type: mRNA

A:Residues: 1030-3002 <MAS>  
A:Cross-references: EMBL:X63556  
R:Oleitz, H.C.; Valle, D.; Francomano, C.A.; Kendzior, R.J.  
Science 259, 680-683, 1993  
A:Title: The skipping of constitutive exons in vivo induced by nonsense mutations.  
A:Reference number: I59574; MUID:93157831  
A:Accession: I59574  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 2217-2288 'T', 2290-2325 <RES>  
A:Cross-references: GB:S54426; NID:9264860; PIDN:AAB2244.1; PID:9264861  
R:Lee, B.; Godfrey, M.; Vitale, E.; Horl, H.; Matzel, M.G.; Sartarazi, M.; Tsipouras, P.  
Nature 352, 330-334, 1991  
A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two different reference number: S17062; MUID:91304567  
A:Accession: S17062  
A:Molecule type: mRNA  
A:Residues: VLVTVVTFUSYNKM', 944-1444 <LEFI>  
A:Cross-references: EMBL:X62008; NID:931398; PIDN:CAB56534.1; PID:95924015  
A:Accession: S62111  
A:Molecule type: protein  
A:Residues: 1166-1176 'X', 1178-1180 'D', 1182-1185 <LE2>  
R:Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Gnanville, R.W.  
J. Biol. Chem. 264, 21381-21385, 1989  
A:Title: Connective tissue microfibrils. Isolation and characterization of three large F  
A:Reference number: A34198; MUID:90078246  
A:Accession: A34198  
A:Molecule type: protein  
A:Residues: 565-575;1890-1892, 'T', 1894-1900 <MAD>  
C:Comment: Fibrillin is a major component of elastin-associated microfibrils.  
C:Genetics:  
A:Gene: GDB:FBNI  
A:Cross-references: GDB:127115; OMIM:134797; OMIM:154700  
A:Map position: 15q21.1-15q21.1  
A:Introns: 2236/1; 2258/1, 2297/1  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein; M  
F:13002/Product: fibrillin (5'-region exon A splice form) (fragment) #status predicted  
F:1332-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MANTC>  
F:1457-1482/Domain: EGF homology <EGF>  
F:2262-2295/Domain: EGF homology <EGF1>

Query Match 13.4%; Score 268.5; DB 2; Length 3002;  
Best Local Similarity 25.6%; Pred. No. 1.6e-09;  
Matches 104; Conservative 31; Mismatches 122; Indels 149; Gaps 24;

QY 30 PCHRCGLVDFK---NQGVDTAKNFGGNTAMEKTLSESEIRLEILEGLCES 85  
DB 1298 PNGRCVNLICKYOCACNPGYHSTPDRLF-----CVIDECSINNGCET 1341  
QY 86 -----SDFECN-----QMLEAOEHLKSEYPRDLFFEMFYCYTLKVC-----CS- 129  
DB 1342 FCTNSEGSYECSCOPGFALMPDR-----SCTDIDE--CEDNPICGGGCTN 1387  
QY 130 -PGYFGPCDLACOG-----GSGRPCSGNGHC-----SGD-GSRQDGGSCRGHGYGP 175  
DB 1388 IPGEY--KCLCYDFMASEDMKTCVDVNECDLNPICLSTGTCENTKSGFICHOMGYSK 1445  
QY 176 L---CTDCMDGYFSSLRNETHSICTACDESKTCSGLTNRDCECEVGVWLDGACVDV 231  
DB 1446 KGKGTCTDINCEIGAHNCGHAYCTNMGSK-CS-----CSGWTGGDGIKCTDL 1495  
QY 232 DECAAEPPPSAOFCKNANGSYTC-----EEC-----DSSCYGCT 267  
DB 1496 DECSNGTHMCSQHADCKNTMGSYCLCKEGYTGDFCTDLDCESENLTGCGNQCINAP 1555  
QY 268 G-----EGP-----GNCK-----ECISGYA--REHGQ 287  
DB 1556 GGYNCEDCMGFVPSADGKACEDIDECSLPNTCVGTCHNLPLGRCECEIETELDRSGN 1615  
QY 288 CADVDECSLAECTCVRRNENCYNTPGSYVCVCPDGF--ETEDACV 331

DB 1616 CTDVNEC-LDPTTCT--SGNCVNTPGSYICCCPPDFELNPRVGCV 1658  
RESULT 5  
A55494  
latent transforming growth factor-beta-binding protein - human  
C:Species: Homo sapiens (man)  
C:Date: 03-Mar-1995 #sequence\_revision 03-Mar-1995 #text\_change 11-Jan-2000  
A:Accession: A55494  
R:Koren, A.; Olofsson, A.; Stenman, G.; Sahlin, P.; Kanazaki, T.; Claesson-Welsh, L.;  
J. Biol. Chem. 269, 32469-32478, 1994  
A:Title: Identification and characterization of LTBP-2, a novel latent transforming  
A:Reference number: A55494; MUID:95096101  
A:Accession: A55494  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1820 <MOR>  
A:Cross-references: GB:Z37976  
C:Genetics:  
A:Gene: GDB:LTBP2  
A:Cross-references: GDB:568901  
A:Map position: 11pter-11pter  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
F:1222-1257/Domain: EGF homology <EGF>  
F:1525-1565/Domain: EGF homology <EGF1>

Query Match 13.4%; Score 268; DB 2; Length 1820;  
Best Local Similarity 25.4%; Pred. No. 1.2e-09;  
Matches 90; Conservative 24; Mismatches 94; Indels 146; Gaps 19;

QY 113 PDLEFWCVKTLKCCSFG-----TYGPCLACOGSGSORPCSGNHCSDGSGNG 162  
DB 865 PDGTR--CV-----CSPGYQLHPSQAYCTDNDCL-----RDPCCGKRCT--NRVG 907  
QY 163 DGSRCRHNGY-----OGPLCT-----DCMGGYFSSLRNE 191  
DB 908 SYSFCYCPGYLMTASGATQECQDINECPQVCSGGGCTNREGSHCCDGGITMVRKH 967  
QY 192 THST-----CTACDESKTCSG-----LTNRDCG----- 215  
DB 968 QQDINECRHPGTCPDGRCVNSPGSYTCLACEGGRGSGSCVDVNECLTPGCAHAKCTN 1027  
QY 216 -----ECEYGMWL--DEGACVDVDECAAEPPPCSAOFCNANGSYTCECCSS----- 262  
DB 1028 LEGSFRCSCEGYEVTSDCKQDQVDECAAR--ASCPTG-LCLMTGSGFACSCACENGWYN 1085  
QY 263 -----CVG--CTG-EGPGNCKECISGYARE--HGOCADVDECSLAECT 300  
DB 1086 EDGTACEDLDCAAPGYCPSGVCNTMGASFCKKCDGGRSPPLDGSCEVDVDECEPDSS 1145  
QY 301 CVRRNENCYNTPGSYVCVCPDGFET-----EDACVPAAEAETEG 341  
DB 1146 CL--GGECKNTVGSYQCLCPQGFOLANGTCEDVNECMGEFHCAPHGCLNSHG 1197  
RESULT 6  
S34583  
serine proteinase (EC 3.4.21.-) PC6B - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 05-Nov-1999  
A:Accession: S34583  
R:Nakagawa, T.; Murakami, K.; Nakayama, K.  
FEBS Lett. 327, 165-171, 1993  
A:Title: Identification of an isoform with an extremely large Cys-rich region of PC6.  
A:Reference number: S34583; MUID:93327934  
A:Accession: S34583  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1548 <NAK>  
A:Cross-references: GB:D17583; NID:9407344; PIDN:BAA04507.1; PID:di005033; PID:944037  
C:Keywords: hydrolase; serine proteinase



Db 291 KLOCKSEFIDALGNCIDINECLISAPCPIGHNTINTGYSYTCOK----- 336  
Oy 273 NCKECISGY-AREHG-OCADVDECSLAECTVKRNENYNTPGSYVCPDG--FEETED 328  
Db 337 NVPCGKGYHLNEGTCTRCVDDECAAPAEPC-GKGHRCVNSPGRCECKTGTYFDGTR 395  
Oy 329 ACVPAEAEATEG 341  
Db 396 MCVNNECORRYPG 408  
RESULT 9  
S78040  
fibulin, splice form C precursor - mouse  
N:Alternate names: basement-membrane protein BM-90  
C:Species: Mus musculus (house mouse)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 23-Mar-2001  
C:Accession: S78040; S78560; S36440  
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.  
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.  
submitted to the EMBL Data Library, January 1993  
A:Description: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dep  
F:30-605/Product: fibulin, splice form C #status predicted <MAT>  
F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted  
A:Title: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dependent  
A:Reference number: S34968; MUID:93358897  
A:Accession: S78040  
A:Molecule type: mRNA  
A:Residues: 1-685 <PAN>  
A:Cross-references: EMBL:X70854  
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.  
submitted to the EMBL Data Library, January 1993  
A:Description: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dep  
F:30-605/Product: fibulin, splice form C #status predicted <MAT>  
F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted  
A:Title: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dependent  
A:Reference number: S34968; MUID:93358897  
A:Accession: S78040  
A:Molecule type: mRNA  
A:Residues: 1-39, 'P', '41-685 <CHU>  
A:Cross-references: EMBL:X70854  
C:Genetics:  
A:Insertions: 568/3  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: alternative splicing; basement membrane; calcium binding; extracellular mat  
F:1-29/Domin: signal sequence #status predicted <SIG>  
F:30-605/Product: fibulin, splice form C #status predicted <MAT>  
F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Query Match 12.6%; Score 253.5; DB 2; Length 685;  
Best Local Similarity 27.1%; Pred. No. 4.6e-09;  
Matches 80; Conservative 24; Mismatches 108; Indels 83; Gaps 14;  
Oy 115 LFEMFCVTKLVCCSPGTGYPDCIACOGS--QRP-----CSGNHCSSG 156  
Db 131 MISYCGGLVFACCVKARENSDFVGNAGADLQDPAKIPDEEDQDPLYNDRCGGGPKQ 190  
Oy 157 DGSROGDC-SCRCHMGVY---GPICTD---CMGYSF---SLRNETHSICTACDESC 203  
Db 191 QCRDTGDEVITSCFVGYLOSGVSCEDINECTGSHNCRLESCINTVGSFRCORDSSC 250  
Oy 204 KTCGSLTN---RDCGEV-----GWVLDG-GACVD 230  
Db 251 CTGYELTEDNNCKDIDECETGHNCPDFICONTLGSFRCPKLOCKSGFIDALGNCID 310  
Oy 231 VDECAEPFPCSAOFCRNANGSYTCEDSSCVCTGEGPNCKECISGY-AREHG-QC 288  
Db 311 INECISISAPCPVGTCTINTGYSYTCOK-----NVNCGRGYHLNEGTIRC 356  
Oy 289 ADVDECSLAECTVKRNENYNTPGSYVCPDG--FEETEDACVPAEATEG 341  
Db 357 VDVECAPAEPC-GKGHCLNPSGFRCECKAGFYFDGISTRCVDINECORYPG 410  
RESULT 10  
S34968  
fibulin, splice form D precursor - mouse  
N:Alternate names: basement-membrane protein BM-90; calcium-binding protein BM-90  
C:Species: Mus musculus (house mouse)

C:Date: 10-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 31-Jan-2000  
C:Accession: S34968; S36441; S13814  
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.  
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.  
submitted to the EMBL Data Library, January 1993  
A:Description: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dep  
F:30-705/Product: fibulin, splice form D #status predicted <MAT>  
F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted  
A:Title: Characterization of a novel calcium-binding 90-kDa glycoprotein (BM-90) shar  
A:Reference number: S13814; MUID:91065369  
A:Accession: S13814  
A:Molecule type: protein  
A:Residues: 28:31-49, 'X', '51-53; 'XX', '110-117; '231-240, 'X', '242-243; '339-362, 'S', '364-387; '4  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: alternative splicing; basement membrane; calcium binding; extracellular m  
F:1-29/Domin: signal sequence #status predicted <SIG>  
F:30-705/Product: fibulin, splice form D #status predicted <MAT>  
F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Query Match 12.6%; Score 253.5; DB 2; Length 705;  
Best Local Similarity 27.1%; Pred. No. 4.7e-09;  
Matches 80; Conservative 24; Mismatches 108; Indels 83; Gaps 14;  
Oy 115 LFEMFCVTKLVCCSPGTGYPDCIACOGS--QRP-----CSGNHCSSG 156  
Db 131 MISYCGGLVFACCVKARENSDFVGNAGADLQDPAKIPDEEDQDPLYNDRCGGGPKQ 190  
Oy 157 DGSROGDC-SCRCHMGVY---GPICTD---CMGYSF---SLRNETHSICTACDESC 203  
Db 191 QCRDTGDEVITSCFVGYLOSGVSCEDINECTGSHNCRLESCINTVGSFRCORDSSC 250  
Oy 204 KTCGSLTN---RDCGEV-----GWVLDG-GACVD 230  
Db 251 CTGYELTEDNNCKDIDECETGHNCPDFICONTLGSFRCPKLOCKSGFIDALGNCID 310  
Oy 231 VDECAEPFPCSAOFCRNANGSYTCEDSSCVCTGEGPNCKECISGY-AREHG-QC 288  
Db 311 INECISISAPCPVGTCTINTGYSYTCOK-----NVNCGRGYHLNEGTIRC 356  
Oy 289 ADVDECSLAECTVKRNENYNTPGSYVCPDG--FEETEDACVPAEATEG 341  
Db 357 VDVECAPAEPC-GKGHCLNPSGFRCECKAGFYFDGISTRCVDINECORYPG 410  
RESULT 11  
T34513  
hypothetical protein ZK783.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34513  
R:Favell, A.; Vaudin, M.  
submitted to the EMBL Data Library, August 1994  
A:Description: The sequence of C. elegans cosmid ZK783.  
A:Reference number: Z21536  
A:Accession: T34513  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3507 <FAV>  
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1  
A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:  
 A:Gene: CESP:2K783.1  
 A:Map position: 3  
 A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;  
 3504/1

Query Match 12.6%; Score 253; DB 2; Length 3507;  
 Best Local Similarity 19.7%; Pred. No. 1.8e-08;  
 Matches 103; Conservative 41; Mismatches 100; Indels 280; Gaps 23;

QY 83 CESSDEPCQMLAEQEHLEAWMLQKSEYPDLEFMEVCYKTK----- 125  
 DB 1454 CATGHNCHESARCO-----NYVGYACFCPTGFRKADGSCODIDECTEHN 1500  
 QY 126 -VCC-----SPGTGYGDC-----LACGGGSRPC-----SGNGHSG----- 156  
 DB 1501 STCCGANAKCVKPGTYSCECENGFLGDGYQCVPTTKPCDSTQSKSHCSSENSCEVD 1560  
 QY 157 --DGS-----RQDGSRCRHMGYGP- 175  
 DB 1561 TVDGSVECKEKGKSGKVCEDINECVAEKAPCSLNMVNMGTFCSCCKQGRGDG 1620  
 QY 176 -LCTD-----C--MGYFSSLNETHS-----ICT-----ACDESK 204  
 DB 1621 FMCSTDINECDERHPCHPNAECTNLEB---SFKCECHSGEGDGIRKCTNPLERSCEDEVER 1677  
 QY 205 TCSGLTNRPC-----GCEVGVWLDE--GACVDVECAAEPPPCS-AAQPFCK 248  
 DB 1678 FCGRDVHSCSLVRIYNGSLSSVCECEPFRFEKESNSCVDIDECESRNNDPASAVCV 1737  
 QY 249 NANGSYTCB-----ECDSACYC-----TGEGR 271  
 DB 1738 NTEGSRCECARGYEGEGVCTDIDECRGMAGCDMAMCINRMGSCGCKMAGYTGDA 1797  
 QY 272 -----GNCK----- 275  
 DB 1798 TCIKIEEPRSDKTACTDMSRLCELEKQCTVDEBEVPOGACLPGHHPINTCQSLQI 1857  
 QY 276 -----ECI-----SGYAREHGOCADVECSLAKTCVARKNENCY 309  
 DB 1858 SGLCAQKNCKNKAECIDIHPSHSCSPDFTGDMITDDVDECNNA-GMCDDEMTKE 1916  
 QY 310 NTPGSYVCVPDGFEEETEDACVPAEAETEGESPQLPSREDL 353  
 DB 1917 NTIGSPNVCYLEGFKKVDKCVYDEKKQ-----ENRERI 1950

RESULT 12  
 A55624  
 fibrillin-1 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 11-Jan-2000  
 C:Accession: A55624  
 R.Ylin, W.; Smiley, E.; Germiller, J.; Sanguinetti, C.; Lawton, T.; Pereira, L.; Ramirez, J. Biol. Chem. 270, 1798-1806, 1995  
 A:Title: Primary structure and developmental expression of Fbn-1, the mouse fibrillin 96  
 A:Reference number: A55624; MUID:95130561  
 A:Accession: A55624  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2871 <YIN>  
 A:Cross-references: GB:L29454; NID:9575509; PIDN:AAA56840.1; PID:9575510  
 C:Genetics:  
 A:Gene: Fbn-1  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 F:1201-1236/Domain: EGF Homology <EGF>

Query Match 12.6%; Score 252.5; DB 2; Length 2871;  
 Best Local Similarity 25.1%; Pred. No. 1.6e-08;  
 Matches 102; Conservative 31; Mismatches 124; Indels 149; Gaps 24;

QY 30 PCHRCGLVDKFE-----NQGMVDTAKNFGGNGTAMEEKTLSKYESSEIRLLILEGLCES 85  
 DB 1167 PHGRCVNLIGKYQACANPGYHPTHDLF-----CVDIDECSTIMNGCET 1210  
 QY 86 -----SDECN-----QMLEAQEHLEAWMLQKSEYPDLEFMEVCYKTKC-----CS- 129  
 DB 1211 FCTNSDSEYECSCQPGFALMPDOR-----SCTDIDO--CEDNPNCDGGQCTN 1256  
 QY 130 -PGTYGPDCLACOG-----GSORPCSGNGHC-----SGD-GSRQDGSRCRHMGYGP 175  
 DB 1257 IGEY--RCLCYDGFMASSDMKTCVDVNECDLNPNLCSLTGCEENKRGSTICHCMDGYSK 1314  
 QY 176 L--CTDCMDGYFSSLNETHSICTACDESKTCSGLTNRDGCEGVWLDEBACVDV 231  
 DB 1315 KKGTCGTDINECEIGHNCGRAHVCNTNAGSPK-CS-----CSPGMIDGJICTDL 1364  
 QY 232 DCAAEPPPCSAAPCKNANGSYTC-----BEC-----DSSCYGT 267  
 DB 1365 DECSNGTHMSCSHADCKNTMGYSRLCKDGYTGDEFTCTDDECSBNLMLCGNQCLNAP 1424  
 QY 268 G-----EGP-----GNCK-----ECTSGYA--REHQ 287  
 DB 1425 GGYRCECDMGFPYSADGKACEDIDECSLPNICVFGCHNLPGLFRCECEIGYELDRSGN 1484  
 QY 288 CADVDECSLAETCYARKNENCYNTPGSYVCVPDGFEE-ETEDACV 331  
 DB 1485 CTDVNEC-LDPTTCT--SGNCVNTPGSYTCDSPDELPNTRVGCV 1527

RESULT 13  
 A57278  
 fibrillin-2 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 11-Jan-2000  
 C:Accession: A57278  
 R.Zhang, H.; Hu, W.; Ramirez, F. J. Cell Biol. 129, 1165-1176, 1995  
 A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extrac  
 A:Reference number: A57278; MUID:95263670  
 A:Accession: A57278  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2907 <ZHA>  
 A:Cross-references: GB:L39790; NID:9762830; PIDN:AAA74908.1; PID:9762831  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 F:1239-1274/Domain: EGF homology <EGF>  
 F:2488-2523/Domain: EGF homology <EGF>

Query Match 12.6%; Score 252.5; DB 2; Length 2907;  
 Best Local Similarity 24.8%; Pred. No. 1.6e-08;  
 Matches 103; Conservative 30; Mismatches 123; Indels 159; Gaps 25;

QY 23 EAAKRPTRC-HRCRGLVDKFNQGMVDTAKNNGGNTAMEEKTLSKYESSEIRLLEILEG 81  
 DB 2249 ECAQNPLCAFRCMNTFGSYE-----CTCPVGYG--IREDDKCK-----DLDECAEG 2294  
 QY 82 L--CESSDEPCQMLAEQEHLEAWMLQKSEYPDLEFMEVCYKTKLVCCSPG-TYGPDL 138  
 DB 2295 LHDCEGRGMCKNL-----IGTFMCICTPRGMARRPDE 2327  
 QY 139 ACQGGSQ--RP--CSGNGHCSGDSRQDGSRCRHMGYGPFLC-TDCMDG----- 183  
 DB 2328 GCVDENECHTRKPGICE-NGRCV---NIISYRCECNEBFGQSSSGTECLDNROGLCPAFV 2383  
 QY 184 -----YFSSLN-ETHSICTACD----- 200  
 DB 2384 LQTMQMASSSRNLVTKSEC-CCDGRGNGHQCCELPCTGTAQYKRIKCPHGAYATDGRD 2442  
 QY 201 -ESCKTCSGL-TNROGE-----CEVGVWLDEG--ACVVDCAAEPPPCSAAPFCK 248  
 DB 2443 IDECKMPSLCTNQCQVNTMGSFRCCKKVGYTMDISGTRACVDLDECSOSPCKPCNF--ICK 2500

OY 249 NANGSYTC-----ECDS-----CVGCTGEGPNCCKEISGYAR 283  
Db 2501 NTKGSYCCSPRGVLDGDKTKDLEDCQTKQHNCOFLCNLTG---GFYCKCPPEFTQ 2557  
OY 284 EHGCADVDECSLAETKCVKKNENCTPTGSIYVCVCPDGE-----ETEDAC 330  
Db 2558 HHNACIDNNBCCGSPSLCGAKG-TCQNTPGSFCECGRFLDASGLNCEVDEC 2611

## RESULT 14

T09456  
Intrinsic factor-B12 receptor Cubilin precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
A:Accession: T09456  
R:Kozuyaki, R.; Kristiansen, M.; Silhataroglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N.  
Blood 91, 3593-3600, 1998  
A:Title: The human intrinsic factor-vitamin B12 receptor, cubilin: Molecular characterization.  
A:Reference number: 216677; MUID:98241400  
A:Accession: T09456  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: mRNA  
A:Residues: 1-3623 <KOZ>  
A:Cross-references: EMBL:AF034611; NID:93929528; PIDN:AC82612.1; PID:93929529  
C:Genetics:  
A:Map position: 10p12  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: receptor; vitamin B12 uptake  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-3623/Product: intrinsic factor-B12 receptor #status predicted <MAT>  
F:436-467/Domain: EGF homology <EGF>

Query Match 12.58; Score 250; DB 2; Length 3623;  
Best Local Similarity 29.18; Pred. No. 2.8e-08;  
Matches 76; Conservative 26; Mismatches 95; Indels 64; Gaps 16;  
OY 118 WFCYKTLKVCSPCTYPDCLA-----CG--GSGORPCSGNCHGSDSGSCRCRHM 171  
Db 154 FFLC-----CPQWKPPCLCSADVNECEIYSGTFLSCONGTCV---NTMKSISYCHCPPE 204  
OY 172 YOGFLCT---DCMDGYFSSILRNETHSICACDESCKTCGTLRRDGECEVAVL--DE 225  
Db 205 TYGPGQCAKSKYDCEG---SVARCVHGI---CEDLMREQAGEPRYSC-VCDAGMSSNS 257  
OY 226 GAC-VVDDECAAEPPPCSAOFCNANGSYTCCECDSSCVGCTGEGPNCCKEISGYAR 284  
Db 258 PACTLDREDCSFQPCSTLVOCFNTGSPFCGACPTGMQ-----NGYI-- 302  
OY 285 HGOCADVDECSLAETKCVKKNENCTPTGSIYVCVCPDGE-----ETEDAC 329  
Db 303 ---CEDINECEINNGGCSVAPPEVCVNTPGSSHCQACPGTGGGRVCTLTIDCSVSN 359  
OY 330 CVPPAEATEGESPP--TOLP 348  
Db 360 CHPDASCSSTLGLPLCTCLP 380

## RESULT 15

A54105  
Fibrillin-2 precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 17-Nov-2000  
A:Accession: A54105; S17063; S31101  
R:Zhang, H.; Apfelroth, S.D.; Hu, W.; Davis, E.C.; Sanguinetti, C.; Bonadio, J.; Mecham, J.; Cell Biol. 124, 855-863, 1994  
A:Title: Structure and expression of fibrillin-2, a novel microfibrillar component preferentially expressed in the aorta.  
A:Reference number: A54105; MUID:94165150  
A:Accession: A54105  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: mRNA  
A:Residues: 1-2918 <ZHA>

A:Cross-references: GB:003272  
R:Lee, B.; Godfrey, M.; Vitale, E.; Horl, H.; Mattei, M.G.; Sarifrazi, M.; Tajpour, N.  
Nature 352, 330-334, 1991  
A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two diff  
A:Reference number: S17062; MUID:91304567  
A:Accession: S17063  
A:Molecule type: mRNA  
A:Residues: 752-1489, 1791, 'GS', 1794-1796, 'QLI', 1922-1923, 'LD', 1926, 'P', 1928 <LEB>  
A:Cross-references: EMBL:X62009  
R:Milwicz, D.M.  
Submitted to the EMBL Data Library, December 1992

A:Reference number: S31101  
A:Accession: S31101  
A:Molecule type: mRNA  
A:Residues: 752-1407, 'R', 1409-1489, 1791, 'GS', 1794-1796, 'QLI', 1922-1923, 'LD', 1926, 'P',  
A:Cross-references: EMBL:X62009  
C:Genetics:  
A:Gene: GDB:FBN2  
A:Cross-references: GDB:128122; OMIM:121050  
A:Map position: 5q23-5q31  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: extracellular protein  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-2918/Product: fibrillin-2 #status predicted <MAT>  
F:1245-1280/Domain: EGF homology <EGF>  
F:1970-2013/Domain: EGF homology <EGF>

Query Match 12.48; Score 249.5; DB 2; Length 2918;  
Best Local Similarity 21.18; Pred. No. 2.6e-08;  
Matches 100; Conservative 53; Mismatches 125; Indels 197; Gaps 26;

OY 33 RCRLVDFK-----NQGM-VDTAKN-----FGGG---NR--AAEETLSKYE 69  
Db 1087 KCRNTTSFCKRCNSGALDMEERNCTDIDECRISPLDGGSGICVNTPGSECCFEYE 1146  
OY 70 SSEIRLLEIE-----GLCESSD--PEC-----NOMLEAOEHLLEAMLOLK 109  
Db 1147 SGFMAMNKMIDICERNPLRCRGTCVNTGSGFQCCPLGHELSPSDECDINECLTS 1206  
OY 110 SEYDPLEFWFC-----VTKLVCCSPGTG--PD-----CLACQGGSGORPCSGN 151  
Db 1207 DN-----LCRNKCVMICTYQCSNPGYQATPRDGCCTDIDECMIMNGCCTQCTNS 1259  
OY 152 CHCGSDSGRODSCRCRMG-----GGLPCTD-----CM 181  
Db 1260 -----EGSYECSSSEGYALMPDRSCADIDECENNPIDCGGQCTNPGEYRCLCY 1310  
OY 182 DGYFSSILRNETHSICACDESCKTC-----SGLTN--RDCGEC 217  
Db 1311 DGFAMSHDMKICIDVNECDLNSNMGCECENTGSPFICHOGLGYVKKGTTCGDVDEC 1370  
OY 218 EV-----GWLDEGACVDVDECAAEPPPCSAOFCNANGS 253  
Db 1371 EIGAHCMDMHASCLNIPGSEFCRREGWIGIKICIDIDECSSNGTSCINACVNTPGS 1430  
OY 254 YTCEDSSSCVCTGEG-----PCKCK-ICISGY--AREHG 286  
Db 1431 YRC-ACSE---GFTGDGFTGSDVDECAENILCENGCLNPGAVRCCEBEGFTPASR 1486  
OY 287 QCADVDECSLAETKCVKKNENCTPTGSIYVCVCPDGEETE-----DACVPP 333  
Db 1487 SCQIDDECSF-QNICV---SGTCNNLPGMFHICIDCGYELDTGTGNCIDIDECAP 1538

## RESULT 16

A48225  
Subtilisin-like proprotein convertase (PC 3.4.21.-) PC5 precursor - mouse  
N:Alternate names: kexin homology; serine proteinase PC6  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
A:Accession: A48225; JX0248  
R:Lusson, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.



Query Match	12.4%	Score 248;	DB 1;	length 915;
Best Local Similarity	30.0%	Pred. No. 1.3e+08;		
Matches	66;	Conservative	20;	Mismatches 86;
			Indels	48;
			Gaps	13;

```
Oy      297 A-----EKTQVRNENCY--NTPGSYVC-VPDPGF 323  
          : | | | : | | | | | |  
Db      823 SYLDHSEEGYKSCRCRDNSLTICNGGFKNCSSCPBGY 862
```

A;Molecule type: mRNA

F;26-85/Domain: GLa domain homology <GLA>

```
F;42-6/5/Product: plasma protein S #status predicted <MAT>
F;88-116/Domain: thrombin-sensitive #status predicted <THS>
```

F;121154/Domain: EGF homology <EG1>  
F;161-199/Domain: EGF homology <EG2>  
F;205-241/Domain: EGF homology <EG3>

Query Match	12.3%	Score 247;	DB 1;	Length 675;
Best Local Similarly	26.9%	Pred. No. 1.2e-08;		
Matches 93; Conservative	34;	Mismatches 113;	Indels 106;	Gaps 21.

```

OY      287  --QCADVDECSLAECTVRKNENCYNTPGSYVC--PDGFEETED 328
          | | | | | | | | : | | | | | | | | : |
Db      238  SKSKCDVDECS--ENTCA---QLCVNYPGYSYCYCDGKKGFKLAQD 278

```

## RESULT 18

A;Status: preliminary

A: Molecule type: mRNA  
A: Residues: 1-1221 <PAN>  
A: Cross-references: GB:W55285; NID:q437046; PIDN:CA53040.1; PID:q437047  
R: Saeki, T.; Mann, K.; Murphy, G.; Chu, M.L.; Timpl, R.  
R: J. Biochem. 240: 427-434, 1996

Enl.: 0, Biochem. 240, 427-438, 1950  
A:Title: Different susceptibilities of fibulin-1 and fibulin-2 to cleavage by matrix  
A:Reference number: S74094, MUID:96435073

A:Accession: S74095  
A:Molecule type: protein  
A:Residues: 236-238, 'X', 240-247; 260-275; 336-344, 'L', 346-361; 405-426; 566-568, 'EM', 569-570  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: calcium binding; duplication; extracellular matrix; glycoprotein; homotrimer  
F:942-978/Domain: EGF homology <EGF>

Query Match	12.3%	Score 247	DB 2	length 1221
Best Local Similarity	22.1%	Pred. No. 1.9e-08		
Matches 94	Conservative 30	Mismatches 132	Indels 170	Gaps 17

```

0Y      83 CESS--DFECNOML-----EAOEHLPAWMLQKLS 110
      ||:      ||:      ||:      ||:      ||:

```



```
D0 525 CESNPINIGYPCONHYMLSCBGGEBLIVPVRBPPEAPRVRVSEMMASRALSLJGTEA 584
QY 111 EYP-----DLFEMFCVKTL--KVCCSPGY---GPDCIACOGGSQ-- 145
D0 585 ELPSLPBGDDODECLMPLGELCQLINTWGSYRCACFPGEFLQGDBORTCRPDGRANPQLD 644
QY 146 -----RPCSGNGHCSGDSRGCD-GSCRCHMGY----QG 174
D0 645 TARESAPRESASAQVSPNTIPLPVPQPTCKDNGPCROCVRYVDGTAMASCPEFYALIMADG 704
QY 175 PLCTID-----CMDGYSSLSLNETHST----- 195
D0 705 VSCDEDOECILMTGHDCSMKQFCVNTLLGSFYCANHTVLCAEGLILNAHRKCVDINDCEVTDL 764
QY 196 --CLACDESKCTGGGLNRRCGBEVSWYLDEGACVDVDEBCAAEPFCSAODFCKNANGS 253
D0 765 HTCTRACHCVNTPGSFOCYKALICEPGYVLTLDGETVDDECYVGNHCQAGFSQNTKGS 824
QY 254 YTCCECSSVCVGCGEBEFGNC---KEIS-----GY 281
D0 825 FYCCARORCMDFLOQDEGMCVDINETSTLLEPCRSGFSCINTWGSYTQCRNPLVLCGRY 884
QY 282 -AREHG-OCADVDECSLAERTCYARKNNENCNTPBSYVCPRDGEFEETEDA----CYPPAE 335
D0 885 HANEESGEADVNECEMGVHNR-GEQOLCYNLPGSYRKCDCKPQFO--RDAFGRTCIDVNE 941
QY 336 AEATEEG 341
D0 942 CWSVPG 947
```

RESULT 19

A57293

latent transforming growth factor beta-binding protein 3 precursor - mouse

N:Alternate names: mitosis-inhibitory peptide

C:Species: Mus musculus (house mouse)

C:Date: 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 11-Jan-2000

C:Accession: A57293; A60487

R:Yin, W.; Smiley, E.; Gernliller, J.; Mecham, R.P.; Florer, J.B.; Wensstrupp, R.J.; Bonadad

J. Biol. Chem. 270, 10147-10160, 1995

A:Title: Isolation of a novel latent transforming growth factor beta binding protein gene

A:Reference number: A57293; MUID:95247723

A:Accession: A57293

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1251 <YIN>

A:Cross-references: GB:L40459

R:Reichelt, K.L.; Paulsen, J.F.; Eljjo, K.

Virchows Arch. B Cell Pathol. 59, 137-142, 1990

A:Title: Isolation of a growth and mitosis inhibitory peptide from mouse liver.

A:Reference number: A60487

A:Accession: A60487

A:Molecule type: protein

A:Residues: 65-69 <REI>

C:Comment: The molecular source of this pentapeptide has not been shown but it correspor

C:Genetics:

A:Gene: Ltbp-3

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: liver; pyroglyutamic acid

F:338-373/Domain: EGF homology <EGF>

F:65/Modified site: pyrrolidone carboxylic acid (Gln) (In mature form) #status experim

```

Db      564  CGHGCGVGPSPDYSCHCHCAAGYRSHRQHR-----YGVVDVNE--CEAEPC 604
QY      134  GPDLCLAO--GGSORPCSGNGHCS-----GDGSRQ-----GDG----- 164
           ||| |||
Db      605  GPGKGICMNTGGSY-----NCHCRNGYRLHVGAGRGSCVDLNECAKPHLLGDGGFCINP 659
QY      165  ---SCRHMGY-----GGPILTD-----CMDDGFSSLRNETHSI--CTADESKKTC 206
           ||| ||| ||| ||| ||| ||| ||| |||
Db      660  GHYKNCNCPGYRLKASRPILTEDIDECRDPSTCPDG---KCEKNPGSEFKCIACOPGYSQ 716
QY      207  SGLFRNDCGECEVGMVLDEGACVVDDECAAPREPCSAOFCNKANNGSYTC----- 256
           ||| ||| ||| ||| ||| ||| ||| |||
Db      717  GG-----GACRYVNEC--SGEPTCSYG--MCEKLPSTYRCTCAAGIRTRT 757
QY      257  -----EECDSSCV---GCTGEGPGNCK--ECISGY--AREHGQCADVDECSLAEK-- 299
           ||| ||| ||| ||| ||| ||| ||| |||
Db      758  GRLSCIDVDDECAENGVCDGIGTTPSPSQCGCLSGYHLSDRSRCEIDEDCPRACIG 817
QY      300  ---TCYRKNKN-----CYTPPSYCYVCYCPDGF 323
           ||| ||| ||| ||| ||| ||| ||| |||
Db      818  GDCLITPNSYRCLCPLAGHRLVGRKKCKKIDEDCSQDGLCLPHACENMIGSYVCVDEGF
QY      324  EETED 328
           |||
Db      878  TLITQD 882

```

RESULT 20  
T22812  
hypothetical protein ZC116.3 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T22812; T27494  
R:Burton, J.  
submitted to the EMBL Data Library, June 1996  
A:Accession: T22812  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3871 <MTL>  
A:Cross-references: EMBL:Z74473; PIDN:CAA98952.1; GSPDB:GN00023; CESP:ZC116.3  
A:Experimental source: clone F56H9  
R:Smyle, R.  
submitted to the EMBL Data Library, June 1996  
A:Reference number: Z20376  
A:Accession: T27494  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3871 <MT2>  
A:Cross-references: EMBL:Z74046; PIDN:CAA98557.1; GSPDB:GN00023; CESP:ZC116.3  
A:Experimental source: clone ZC116  
C:Genetics:  
A:Gene: CESP:ZC116.3  
A:Map position: 5  
A:Introns: 29/3; 66/2; 244/1; 359/3; 422/1; 549/1; 572/1; 728/1; 820/1; 889/3;  
2763/1; 2809/1; 2806/3; 3051/3; 3176/1; 3254/1; 3341/2; 3536/1; 3565/2; 3582/3; 3609/1

Oy	270	---	GFNCKKECISGVARRHGOCADVDEC-----SLAKTCYVRKNKNCYNTPGSYYCCP	320
			:	
Db	270	NTLGSIYDCSDSCPTGSGGGCGCNVKRDSVCVKKKHCHLA--TC-KVTDDGTSAVGDTYTCTCP	326	
Oy	321	DGEFET---	EDACVP	334
		:	:	
Db	327	DGYVDGIGEEGCYKSA	343	
RESULT	21			
A55184		fibulin-2 precursor - human		
N:	Alternate names:	protein DKFPz586A1519..1		
C:	Species:	Homo sapiens (man)		
C:	Date:	27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000		
C:	Accession:	A55184; T08744		
R:	Zhang, R.Z.; Pan, T.C.; Zhang, Z.Y.; Mattei, M.G.; Timpl, R.; Chu, M.L.			
G:	Genomics	22, 425-430, 1994		
A:	Title:	Fibulin-2 (FBLIN2) : human cDNA sequence, mRNA expression, and mapping of the gene		
A:	Reference number:	A55184; MUID:95104855		
A:	Accession:	A55184		
A:	Status:	preliminary		
A:	Molecule type:	mRNA		
A:	Residues:	1-1184 <MA>		
A:	Cross-references:	(GB:X82494; NID:g575232; PIDN:CAA57876.1; PID:g575233		
R:	Mambuit, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.			
A:	Submitted to the Protein Sequence Database,	March 1999		
A:	Reference number:	T216471		
A:	Accession:	T08744		
A:	Molecule type:	mRNA		
A:	Residues:	656-719, /ODEICMGAMHDCSRQFCVNLTGSPFYCVNHFTVLICADGYILNAHRKCVD', /720-853, 'T', /855-1118		
A:	Cross-references:	EMBL:AL050095		
A:	Experimental source:	adult uterus; clone DKFPz586A1519		
C:	Genetics:			
A:	Gene:	GDB:FBLIN2		
A:	Cross-references:	GDB:293037; OMIM:135821		
A:	Map position:	3p25-3p24		
A:	Note:	DKFPz586A1519..1		
C:	Superfamily:	unassigned EGF-related proteins; EGF homology		
C:	Keywords:	alternative splicing; extracellular matrix		
F:	1-37/Domain:	signal sequence #status predicted <SIG>		
F:	28-1184/Product:	fibulin-2 protein #status predicted <MAT>		
F:	905-941/Domain:	EGF homology <EGF>		
Query Match	12.1%;	Score 242.5;	DB 2;	Length 1184;
Best Local Similarity	23.8%;	Pred. No. 3.5e-08;		
Matches	88;	Conservative .38;	Mismatches 115;	Indels 129;
			Gaps	19;
Oy	83	CESS---DFECNQML-	EAQEHLEAWWLQKS	110
	:	:	:	:
Db	535	CESNNLGYPCNHVALSCOEGLIPEVRPRPPAPAAARVSAEAGRAALSIGTEA	594	
Oy	111	EYP-----DLEFWCYKTL--KYCCSPG-TYGPDCIAC-QGSSQRP-	147	
	:	: :	:	:
Db	595	ELPNLSLPDGDDECELLRLPGELQLHCINTVSGSYHCACEPFESLDODGRCTRPEGHHPQE	654	
Oy	148	-----CSGNHGCSGGSSRGDGS-CRCNHGX-----QG	174	
	:	:	:	:
Db	655	AQPQALSKSEFSQVASNITPLPLPQPNTRCKDNKGPKQYCSITYGSAITSCEPFGVIAMDG	714	
Oy	175	PLCTD---CMDGFSSLRNETHSICTACDESCRTCSGLTNRDQCEBEGWVLDGACVDY	231	
	:	:	:	:
Db	715	VSCDEINECVTFDLTHCSKGE-HCVNTLGSSFHC-YKALT-----CEPGYALKDGCEBDY	765	
Oy	232	DECAAEPPPCSAQAQCCKANANGSYTCEECDSSCVGCTGEGPNGCKKCIIGIYARE-HGQCAD	290	
	:	:	:	:
Db	766	DECAAGTHTCOPGFLCONTKTSFYQA-----RORCMDGFILODPDEGNCVD	810	
Oy	291	VDECSIAKTCYVRKNKNCYNTPGSYYC-----VCPDGEETEEDA--CVPPAAE-----A	338	
	:::::	:::::	:::::	:::::
Db	811	INECTSLSEPC-RPFSCINTIVGSTTCQRNPILICARGIHASDGAACYVDVNECFEGVIRC	869	

[illegible]

A:Title: Intron-exon organization of the active human protein S gene Psalpa and its psc  
 A:Reference number: A35611; MUID:91084445  
 A:Accession: A35611  
 A:Molecule type: DNA  
 A:Residues: 1-25 <PL3>  
 A:Cross-references: GB:J02918  
 R:Hoskins, J.; Norman, D.K.; Beckmann, R.J.; Long, G.L.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 349-353, 1987  
 A:Title: Cloning and characterization of human liver cDNA encoding a protein S precursor  
 A:Reference number: A26157; MUID:87092407  
 A:Accession: A26157  
 A:Molecule type: mRNA  
 A:Residues: 1-10, 'P', 12-25, 'L', 27-676 <HOS>  
 A:Cross-references: GB:M15036; NID:9190288; PIDN:AAA36479.1; PID:9190289  
 R:Lundvall, A.; Dackowski, W.; Cohen, E.; Shaffer, M.; Mahr, A.; Dahlbeck, B.; Stenflo, Proc. Natl. Acad. Sci. U.S.A. 83, 6716-6720, 1986  
 A:Title: Isolation and sequence of the cDNA for human protein S, a regulator of blood coagulation  
 A:Reference number: A25891; MUID:86313649  
 A:Accession: A25891  
 A:Molecule type: mRNA  
 A:Residues: 27-220, 'L', 222-262, 'H', 264-344, 'Y', 346-676 <LUN>  
 A:Cross-references: GB:M14338; NID:9190448; PIDN:AAA60181.1; PID:9190449  
 A:Note: part of this sequence, including the amino end of the mature protein, was determined by Randerud, C.M.; Lundvall, A.; Wydro, R.; Stenflo, J.  
 Biochemistry 29, 7861-7868, 1990  
 A:Title: Molecular analysis of the gene for vitamin K dependent protein S and its pseudogene  
 A:Reference number: A35612; MUID:91084446  
 A:Accession: A35612  
 A>Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 284-676 <EDE>  
 A:Cross-references: GB:J02919  
 R:Ploos van Amstel, J.K.; Van der Zanden, A.L.; Bakker, E.; Reitsma, P.H.; Bertina, R.M.; Thromb. Haemost. 58, 987-987, 1987  
 A:Title: Two genes homologous with human protein S cDNA are located on chromosome 3.  
 A:Reference number: A60903; MUID:88178564  
 A:Accession: A60903  
 A:Molecule type: mRNA  
 A:Residues: 351-676 <PLO>  
 R:Ploos van Amstel, J.K.; Van der Zanden, A.L.; Reitsma, P.H.; Bertina, R.M.; FEBS Lett. 222, 186-190, 1987  
 A:Title: Human protein S cDNA encodes Phe-16 and Tyr 222 in consensus sequences for the A:Reference number: S02424; MUID:88005138  
 A:Accession: S02424  
 A:Molecule type: mRNA  
 A:Residues: 1-676 <PL2>  
 A:Cross-references: EMBL:Y00692; NID:936578; PIDN:CAA68667.1; PID:936579  
 C:Genetics:  
 A:Gene: GDB:PROS1; PROS  
 A:Cross-references: GDB:120721; OMIM:176880  
 A:Map position: 3p11.1-3q11.2  
 A:Insertions: 26/1; 78/3; 87/1; 116/1; 157/1; 201/1; 243/1; 283/3; 322/2; 385/3; 441/3; 498/3  
 C:Complex: in plasma forms a complex with Cb binding protein  
 C:Function:  
 A:Description: a cofactor for activated protein C (EC 3.4.21.69); thrombin cleavage dest  
 C:Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat hom  
 C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; blood coagulation; carbox  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-41/Domain: propeptide #status predicted <PRO>  
 F:26-85/Domain: Gla domain homology <GLA>  
 F:42-676/Product: plasma protein S #status predicted <MAT>  
 F:121-154/Domain: EGF homology <EG1>  
 F:161-199/Domain: EGF homology <EG2>  
 F:205-241/Domain: EGF homology <EG3>  
 F:247-282/Domain: EGF homology <EG4>  
 F:315-667/Domain: sex hormone-binding globulin homology <SHB>  
 F:335-678/Domain: laminin G repeat homology <LGR>  
 F:47,48,55,57,60,61,66,67,70,73,77/Modified site: gamma-carboxylglutamic acid (Glu) #stat  
 F:58-63,88-113,121-134,126-143,145-154,161-175,171-184,186-199,205-217,212-226,228-241,2  
 F:111-112/Cleavage site: Arg-Ser (thrombin) #status predicted  
 F:136/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted  
 F:177,219,258/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
 F:499,509,530/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 11.9% Score 239; DB 1; Length 676;  
 Best Local Similarity 25.7% Pred. No. 3.8e-08;  
 Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;  
 QY 11 LPLLLLPAPAEAKKPTPCRCRGLVDFNQGQVDTAKKFGGNTAMEKTLTKYES 70  
 DB 11 LLACLLLVLPVSEA-----NLSKQQAQVLYRRKR--ANSLEETQGNLER 56  
 QY 71 SEIRLEITLLEGLCESSDEPCNMLE--AQEHLBAMVL-----OLKSEYP 113  
 DB 57 -----ECIEELCKNKE--EAREVEFENDPETYFPKYLCLURSFOTGLFTAAQSTNAYP 108  
 QY 114 DLFFMPCVKTLKVCSSGTYTPDGLAOGGSGRSCSGHCS-GDGSNGSGSCRCMHGY 172  
 DB 109 DLRS--CVNALPDCSP-----LPCNDEGYSCKDG--KASFTCTCKPGW 149  
 QY 173 QGPLC---TDCMDGYFSSLNETHSICTACDESKCTCSGLTNRDCEGEVWLV--DEG 226  
 DB 150 QGEKCEPDINCKD-----PSNINGSQSQICDNTPG--SYHC--SKCKGFVMLSKK 197  
 QY 227 ACVDVDECAEPPEPCSAQAQCKNANGSYTCECDSSCGVGTGEGPGNCKECISGYA--RE 284  
 DB 198 DCKDVDECSLKPISICGTA-VCKNIPGDFEC-----ECPEGYRYWLK 237  
 QY 285 HGQCADVDECSLAKTCVARKNENYNTPGSYVCV--PDGEFETD-----ACVP 332  
 DB 238 SKSCEDIKES--ENMCA--QLCVNPGGYTCYCDCKKGRKLAQDKSCFVSVCLP 290  
 RESULT 24  
 B48225  
 Probable proprotein convertase (EC 3.4.21.-) 5 precursor - rat  
 N:Alternate names: PCS precursor  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 23-Feb-1997  
 C:Accession: B48225  
 R:Lusson, J.; Vieau, D.; Hemelijn, J.; Day, R.; Chretien, M.; Seldah, N.G.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993  
 A:Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate p  
 A:Reference number: A48225; MUID:93342056  
 A:Accession: B48225  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-915 <LUS>  
 A:Cross-references: GB:L14933  
 C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology  
 C:Keywords: duplication; glycoprotein; hydrolyase; integrin binding; serine proteinase  
 F:1-34/Domain: signal sequence #status predicted <SIG>  
 F:35-915/Product: probable proprotein convertase 5 #status predicted <PRO>  
 F:117-915/Product: probable proprotein convertase 5 #status experimental <MAT>  
 F:164-402/Domain: subtilisin homology <SBT>  
 F:173,214,388/Active site: Asp, His, Ser #status predicted

Query Match 11.9% Score 239; DB 2; Length 915;  
 Best Local Similarity 28.6% Pred. No. 4.8e-08;  
 Matches 66; Conservative 16; Mismatches 79; Indels 70; Gaps 12;  
 QY 125 KVC---GSPGYGPDCLACOGSGRPGSGNGHSGDGSROGDSRCRHMGYQGPLCTDCM 181  
 DB 670 RICYSSCPGPHFADKKRC-----RKCAPN-----CESFGSHAOQCLSK 710  
 QY 182 DGYFSSSLNETHS-----ICTACDESKCTCSGLTNRDCEGEVWVLDE 225  
 DB 711 YGYF--LNEETSSCAVACPEGSYODIRKKNICGKSEMCKTCTGFHN--CTECKGGLSLDG 766  
 QY 226 GACVDVDECAEPPEPCSAQAQCKNANGSYTCECDSSCGVGTGEGPGNCKECISGYAREH 285  
 DB 767 SRC-----SVTCEDGQF--SCHDCQPHRCATCAGAGADCCINCTEGVME 813  
 QY 286 GQCADVDECSLA-----EKTCAVARKNENY--NTPGSYVC--VCPDGF 323



Db 767 SRCVSCEDGRYFNGDCQPCCHRECATCAGAGADGCINTCEGYMEDGRVQ----- 818  
QY 239 PPSAQAQFCNNA--NGSYTCEEDSSCVGCTGEGPNCCKECISYAHEHQ-----CAD 290  
Db 819 -SCSISYFPHSSSENGKSKCKDISCLTNGCPFKNCTSCPSGYLLDLGCOMGAICKD 877  
QY 291 VDECSLAEK---TCVRKNENCYMTPGSYVC 317  
Db 878 ATEESMAEGGFCMLYKKNKNCQKRYLQQLC 907

RESULT 28  
T18739  
hypothetical protein B0393.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18739  
R:Stulston, J.  
Submitted to the EMBL Data Library, September 1994  
A:Reference number: Z19013  
A:Accession: T18739  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1106 <MIL>  
A:Cross-references: EMBL:Z17983; NID:el519039; PIDN:CAA86058.1; GSPDB:GN00021; CESP:B039  
A:Experimental source: clone B0393  
C:Genetics:  
A:Gene: CESP:B0393.5  
A:Map position: 3  
A:Introns: 15/3; 60/1; 100/3; 343/2; 443/1; 507/3; 609/3; 659/1; 796/1; 876/1; 946/1; 99

Query Match 11.7% Score 234; DB 2; Length 1106;  
Best Local Similarity 22.2% Pred. No. 1.2e-07;  
Matches 84; Conservative 26; Mismatches 100; Indels 168; Gaps 14;

QY 108 LKSEYPLD---FEW-FCVTKLV-CCSPGYGPDCLACGGGSG--RPGSGNGHC-SGDGS 159  
Db 156 LLSKSDIGIPGEWLFVDDDRITLCGAGFKGLEICSCASQWPNCKSKCHDGDGD 215  
QY 160 RQGDGSC--RCHMGYOG-PLCTDCCMD-----GYFSSL----- 188  
Db 216 DOENGRCPNKCSGPGWIGEBICDEMDCEMGIDNCPNEQPDCLNTPGSFLCLCFEYDEA 275  
QY 189 -----RNEHNS----- 194  
Db 276 QQCKNSKAAPPSAPIDVVIIMHPTFRKPTPPKVTAPPRNRRKSTTAATSPFITETT 335  
QY 195 -----ICTACDESCKTCSCGLTNRDGECE 218  
Db 336 TRQTTTTRKESITPLLTTEVYTRPSPTTYVTAPACARCDQAKNSNGVCT-----CS 390  
QY 219 VGWVLDGACVDVDECAEPSPCSAQAQFCNANGSYTCEECDSVCGCTGEGPNCCKECI 278  
Db 391 EGFGLDGFRCYDVECEIPAVCRDHSICSNITISFEC-----TCH 431  
QY 279 SGYAREHGACADVDECSLAKTC--VRKNENCYMTPGSYVCVCPDGE----- 324  
Db 432 GGRFEEGKCEKDECELEPRICGDPNKGKTCINKKDTFELCLKDGEGDPSSCRDQVNE 491  
QY 325 -ETEDACVPAAEAATEG 341  
Db 492 CKNDPACGPNSSQCTNTG 509

RESULT 29  
KXMS  
plasma protein S precursor - mouse  
N:Alternate names: vitamin K-dependent glycoprotein S  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 26-Jan-1996 #text\_change 16-Jul-1999  
C:Accession: S43504; I59616; S35962

R:Chu, M.D.; Sun, J.; Bird, P.  
Biochim. Biophys. Acta 1217, 325-328, 1994  
A:Title: Cloning and sequencing of a cDNA encoding the murine vitamin K-dependent pro  
A:Reference number: S43504; MUID:94198297  
A:Accession: S43504  
A:Molecule type: mRNA  
A:Residues: 1-675 <CHU>  
A:Cross-references: EMBL:Z25469; NID:9396426; PIDN:CAA80961.1; PID:9396427  
R:Lu, D.; Schmidt, D.K.; Long, G.L.  
Thromb. Res. 74, 135-142, 1994  
A:Title: Structure of mouse protein S as determined by PCR amplification and DNA sequ  
A:Reference number: I59616; MUID:94302659  
A:Accession: I59616  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 35-492, V, 494-675 <RES>  
A:Cross-references: GB:I27439; NID:9487866; PIDN:AAA40006.1; PID:9487867  
C:Complex: in plasma forms a complex with C4b binding protein  
C:Function:  
A:Description: a cofactor for activated protein C (EC 3.4.21.69); thrombin cleavage p  
C:Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat  
C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; blood coagulation; car  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-41/Domain: propeptide #status predicted <PRO>  
F:42-65/Domain: Gla domain homology <GLA>  
F:47-675/Product: plasma protein S #status predicted <MAT>  
F:121-154/Domain: EGF homology <EG1>  
F:161-199/Domain: EGF homology <EG2>  
F:205-241/Domain: EGF homology <EG3>  
F:247-282/Domain: EGF homology <EG4>  
F:315-666/Domain: sex hormone-binding globulin homology <SHB>  
F:325-478/Domain: laminin G repeat homology <LGR>  
F:47-48, 55-57, 60-61, 66-67, 70-73, 77/Modified site: gamma-carboxyglutamic acid (Glu) #  
F:56-63, 88-113, 121-134, 126-143, 145-154, 161-175, 171-184, 186-199, 205-217, 212-226, 228-24  
F:111-112/Cleavage site: Arg-Ser (thrombin) #status predicted  
F:136/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted  
F:177, 219/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F:499, 509/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.6% Score 232; DB 1; Length 675;  
Best Local Similarity 26.3% Pred. No. 1.1e-07;  
Matches 91; Conservative 30; Mismatches 119; Indels 106; Gaps 20;

QY 11 LRLPLLLPRAPEAAKPTPCRRGCLVDKFNQGMVDTAKNFGCGTAWKEKLTYES 70  
Db 11 LLACLAIVPVSER-----NFLSKERASQVIVRRKR---ANTLFETKGNLDR 56  
QY 71 SEIRLEILEGLCESSDFECNQMLEAOBEH-----LEAWML-----OLKSEYP 113  
Db 57 -----ECIEELCKME--EAEREVENNPETDYFPKYLICGAFRVGSAFHAQSANAYP 108  
QY 114 DLFEWFCVTKLVCCSPGYGPD-CLACQGGSGRPGSGNGHSGDGSRGDGSRCRHMY 172  
Db 109 DLRS--CVKATISDCCDIPICNEDGYLACQDS-----QAAFTCFCKPGW 149  
QY 173 QGPLC-----IDCMGIFSSILRNETHSICTACDESCKTCSCGLTNRDCEGCEVGMWL--DEG 226  
Db 150 QGDRGQYDVNCKD-----PSNVNNGCSQICDNTPG--SYHC--SCKRGFAMLPKK 197  
QY 227 ACVVDVDECAEPSPCSAQAQFCNANGSYTCEECDSVCGCTGEGPNCCKECISYAREHG 286  
Db 198 DCKDLDECALKPSYCGTA-VKKNIPGDEEC-----ECPPDGYRDPDS 237  
QY 287 --QCADVDECSLAERTCVKRNENCYMTPGSYVC--PDGEETED 328  
Db 238 SKSCQDVDECS--ENMCA---QLCVNFPGGYSYCYCDGKKGFKLAOD 278

RESULT 30  
G00043  
osteonidogen - human  
C:Species: Homo sapiens (man)

C>Date: 13-Mar-1997 #sequence\_revision 13-Mar-1997 #text\_change 05-Nov-1999  
C:Accession: G00043  
R:Ohno, I.  
submitted to GenBank, July 1996  
A:Reference number: H00043  
A:Accession: G00043  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1376 <OH>  
A:Cross-references: GB:D86425; NID:g1449166; PIDN:BA13087.1; PID:g1449167  
C:Superfamily: nidogen; EGF homology: LDL receptor WTD-containing repeat homology: thy  
F:806-843/Domain: EGF homology <EGF1>  
F:853-891/Domain: EGF homology <EGF>  
F:941-1006/Domain: thyroglobulin type I repeat homology <THY1>  
F:1020-1085/Domain: thyroglobulin type I repeat homology <THY2>  
F:1242-1286/Domain: LDL receptor WTD-containing repeat homology <W>

Query Match 11.6%; Score 232; DB 2; Length 1376;  
Best Local Similarity 27.8%; Pred. No. 1.9e-07;  
Matches 63; Conservative 12; Mismatches 74; Indels 78; Gaps 7;

QY 147 PCSGNHCGRGDSRQDSCRCMVGQGLCTDCMDGYFSSLRNETHSICTACDESKTC 206  
DB 763 PCY-----DSHMCCTTARCPG-----TGVDYTC--- 787

QY 207 SGLTNRDCEGEVWLDGACVYDECAEPPPCSAOFCCKNANGSYTCE----- 257  
DB 788 -----ECASGYGGDRNCVDENECATGFRGCPNSVCITLPASRYNCECSGYEFAD 838

QY 258 -----ECDSSCVGCTGEGPNCKE-----CISGYAREHGQCADVDECS 295  
DB 839 DRHTCLITPPANPCDGSHTCAAPACQARCVHNGSTFCACLPYAGDGHQCTDYDECS 898

QY 296 LAEKTGVKRNKENVYRGSYVVCVCPDGFETEDACVPAEAETEE 342  
DB 899 --ENRC-HPAATCYNTPGSFSCMCOPGYGDGQCIPDSTSLTPE 942

RESULT 31  
T42681  
hypothetical protein DKFp434E0321.1 - human (fragments)  
C:Species: Homo sapiens (man)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000  
C:Accession: T42681  
R:Blum, H.; Baursechus, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, November 1999  
A:Reference number: Z22233  
A:Accession: T42681  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-870; 871-1069 <AAA>  
A:Cross-references: EMBL:AL133021  
A:Experimental source: adult testis; clone DKFp434E0321  
C:Note: the cDNA sequence contains a +1 frameshift near codon 870  
C:Genetics:  
A:Note: DKFp434E0321.1

Query Match 11.5%; Score 231.5; DB 2; Length 1069;  
Best Local Similarity 26.7%; Pred. No. 1.6e-07;  
Matches 66; Conservative 23; Mismatches 107; Indels 51; Gaps 11;

QY 127 CCSPTGYGPDCLACGSGRPSGNGSGDGSRODGSRCRHMVGQGLCTDCMDGYFS 186  
DB 504 CCK-GYFGDQCACPGPGPAFCNNRGVCLDQYATGE--CKCTGTGNGTACEKCMWGRG 560

QY 187 SLRNETHSICTACDESKTCGSLTNRDCEGEVWLDGACVYDECAEPPPCSA 244  
DB 561 -----PDLCPGCGSDHGDQDGLITGSGQCLCETGWT--GSPCDTQAVLPAYCTPPCSAH 612

QY 245 QCKNANGSYTCE--ECSSCVGCT-----GEGPNCKE-----CISGYA 282  
DB 245 QCKNANGSYTCE--ECSSCVGCT-----GEGPNCKE-----CISGYA 282

DB 613 ATCKENN--TCECNIDEGDITCTVWDECKQDNGGCAKVARCSOKGTYSQCKGYK 669  
QY 283 REHGQCADVDECSLAEKTGVKRNKENVYRGSYVVCVCPDGFETEDACVPAEAETEE 341  
DB 670 GGHGSTEIDPCADGLNGCGCHEHATCKMTGPKRHCCECKSHVGGILNC----- 718

QY 342 ESPQQLP 348  
DB 719 -EPEQLP 724

RESULT 32  
S53434  
plasma protein S precursor, vitamin K dependent - rhesus macaque (fragment)  
C:Species: Macaca mulatta (rhesus macaque)  
C:Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 16-Jul-1999  
C:Accession: S53434  
R:Greengard, J.S.; Fernandez, J.A.; Radtke, K.P.; Giffin, J.H.  
Biochem. J. 305, 397-403, 1995  
A:Title: Identification of candidate residues for interaction of protein S with C4b  
A:Reference number: S53433; MUID:95134217  
A:Accession: S53434  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-642 <GRE>  
A:Cross-references: EMBL:L31380  
A:Experimental source: tissue type liver  
A:Note: the source is designated as rhesus monkey  
C:Genetics:  
A:Gene: PROS  
C:Superfamily: plasma protein S; EGF homology: Gla domain homology; laminin G repeat  
F:1-51/Domain: Gla domain homology (fragment) <GIA>  
F:1-7/Domain: signal sequence (fragment) #status predicted <SIG>  
F:8-642/Product: plasma protein S #status predicted <MAT>  
F:87-120/Domain: EGF homology <EG1>  
F:127-165/Domain: EGF homology <EG2>  
F:171-207/Domain: EGF homology <EG3>  
F:213-248/Domain: EGF homology <EG4>  
F:281-633/Domain: sex hormone-binding globulin homology <SHB>  
F:291-444/Domain: laminin G repeat homology <LGR>

Query Match 11.4%; Score 228.5; DB 2; Length 642;  
Best Local Similarity 29.0%; Pred. No. 1.7e-07;  
Matches 71; Conservative 24; Mismatches 75; Indels 75; Gaps 16;

QY 107 QLKSEYDLEFNFQVTKLVKCSPTGYGPDCLACGSGRPSGNGHCS-GPGSRQDGS 165  
DB 68 QSTDAIPDLRS--CVNAIPDQCS-----LPENEGYNSCKDG--KASFT 108

QY 166 CRCHMGVQGPLC---TDCMDGYFSSLRNETHSICTACDESKTCGSLTNRDCEGEVW 221  
DB 109 CRCKPGMOGEBREFDINEKD-----PKNINGGCSQIDNNIPG--SYHC-SKSGF 156

QY 222 VL--DEGACVYDECAEPPPCSAOFCCKNANGSYTCECDSSVCYCTGEGPNCKECTIS 279  
DB 157 VMLSNKKDKDYDECSIKPNMGTA--VCKNIPGDEEC-----ECPE 196

QY 280 GYA--REHGQCADVDECSLAEKTGVKRNKENVYRGSYVVC--PDGFETED----- 328  
DB 197 GRYNLSKSCBEDVDECS--ENMCA--OLCVNYPGGTYCYDGKKGFLADQKSCNAV 251

QY 329 -ACVP 332  
DB 252 SWCLP 256

RESULT 33  
A43434  
furin (EC 3.4.21.75) 2 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Mar-2000  
C:Accession: A43434

R:Roedbroek, A.J.; Creemers, J.W.; Pauli, I.G.; Kurzik-Dunke, U.; Rentrop, M.; Gateff, E.  
 J. Biol. Chem. 267, 17208-17215, 1992  
 A>Title: Cloning and functional expression of Dfurin2, a subtilisin-like proprotein proc  
 A:Reference number: A43434; MUID:92381036  
 A:Accession: A43434  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1680 <ROE>  
 A:Cross-references: GB:M94375; NID:q157461; PID:q157462  
 A>Note: sequence extracted from NCBI backbone (NCBIN:111933, NCBI:P.111934)  
 C:Genetics:  
 A:Gene: FlyBase:Fur2  
 A:Cross-references: FlyBase:FBgn0004598  
 C:Superfamily: subtilisin homology  
 C:Keywords: hydrolase; serine proteinase; transmembrane protein  
 F:409-652/Domain: subtilisin homology <SBT>  
 F:418,457,638/Active site: Asp, His, Ser #status predicted

Query Match 11.4%; Score 228; DB 2; Length 1680;  
 Best Local Similarity 22.8%; Pred. No. 3.9e-07;  
 Matches 84; Conservative 31; Mismatches 127; Indels 126; Gaps 19;

OY 31 CHRCGLVDKFNQGVDTAKKNGGNTAMEKTLTK-----YESSEIRLELLEGLCE 84  
 DB 1118 CATCNGPTD---QDCITCRSSRY-----AMQNKCLISCPDGFYADKRLKPCQEG-CK 1168  
 OY 85 --SSDFECNMLEAOEHLEAMHQLKSEYRDLFEWFCVTLKVCSPRTYGPCLACG 142  
 DB 1169 TCTISNGVCS-----ECIDNMTLNRKDK-----CIVSGSEGCSESEF---YSQVE 1209  
 OY 143 GSORPCSGN-GHSCGDSRGDSC-----RCHNGYO-----GPLC----- 177  
 DB 1210 GQCRPCASGSCNG-----PADTSCITSPRRRLLEGRVCVSGCREGFVEAGSLCSPCLH 1265  
 OY 178 -----TDCMDGYFSSLRNETHSITCAPDCSCKTCGLTN 211  
 DB 1266 TCSQCVSRITNCNSKGLQLNGECRTTCADGYSD-----RGICACVYLSCHTCSGPRR 1320  
 OY 212 RDGCECEVGVLDGACVDDECAAEPPPCSAQFCNKANANGSYCECDSSCVGCTEGP 271  
 DB 1321 NCQVQCAAGQOLAGEC-----HPECEPGFY-----KSDECCQKCHHYCKTKCNDAGP 1367  
 OY 272 GNCKECTISGAREHGQCAD-----VDECSLAETCVKRNKENCNTGSGYVCVCPDGFEE 325  
 DB 1368 LACTSCPHSMIDGGLCMECLSSGYVDTSATCTKCHDSRSCRG-PGQFSC----- 1418  
 OY 326 TEDACVPP 333  
 DB 1419 --KGCVPP 1424

RESULT 34  
 A35844  
 Xotch protein - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 12-Oct-1990 #sequence\_revision 12-Oct-1990 #text\_change 13-Aug-1999  
 C:Accession: A35844  
 R:Coffman, C.; Harris, W.; Kintner, C.  
 Science 249, 1438-1441, 1990  
 A>Title: Xotch, the Xenopus homolog of Drosophila notch.  
 A:Reference number: A35844; MUID:90385285  
 A:Accession: A35844  
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A:Molecule type: mRNA  
 A:Residues: 1-2524 <COF>  
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology  
 C:Keywords: transmembrane protein  
 F:146-177/Domain: EGF homology <EGX1>  
 F:184-215/Domain: EGF homology <EGF1>  
 F:222-254/Domain: EGF homology <EGF>  
 F:456-487/Domain: EGF homology <EGX2>  
 F:757-788/Domain: EGF homology <EGF3>

F:1025-1056/Domain: EGF homology <EGX3>  
 F:1924-1956/Domain: ankyrin repeat homology <AN1>  
 F:1957-1989/Domain: ankyrin repeat homology <AN2>  
 F:1991-2023/Domain: ankyrin repeat homology <AN3>  
 F:2024-2056/Domain: ankyrin repeat homology <AN4>  
 F:2057-2089/Domain: ankyrin repeat homology <AN5>

Query Match 11.3%; Score 227; DB 2; Length 2524;  
 Best Local Similarity 23.9%; Pred. No. 6.2e-07;  
 Matches 69; Conservative 20; Mismatches 100; Indels 100; Gaps 14;

OY 126 VC-CSPGTGPDCL-----ACGCGSORPCSGNGHSCGSGSRGDSRCRHMGYGPLCTD- 179  
 DB 165 ICKRPPEFHATCKQODINEC-----SQNPCKNGQGINE---FGSRTGQCNFTGRNDEP 218  
 OY 180 -----CMDGYFSSLRNETHSICTA-----CDESKCTC-----SGLTN 211  
 DB 219 YVPCNPSPCLNGCTGCRQTDITSTYDCTCLPGFSQNCSENIIDCPSNNCRNGCTCVDGVNT 278  
 OY 212 RDGCECEVGVLDGACVDDECAAEPPPCSAQFCNKANANGSYC-----ECCDSS- 262  
 DB 279 YNC-QCPPDWT-GQYCTEDVDECOIMPNACQNGTCHNTYGGYNCVCGMTGDCSENI 336  
 OY 263 --CVGCTGEGPGNCKE-----CISGYAREHGQ----- 288  
 DB 337 DDCANAAHSGATCHDVAISFYCECPHGRIGLLCHLDNACISNPNCSNCDTNPVNGKA 396  
 OY 289 -----ADVDECSLAETCVKRNKENCNTGSGYVCVCPDGF 323  
 DB 397 ICTCPGVTGTPACNNDVDECSLGNPCER-GGRCTNLTGSPQCNCPOGY 444

RESULT 35  
 A46019  
 Notch-1 protein - mouse  
 N:Alternate names: notch protein  
 C:Species: Mus musculus (house mouse)  
 C>Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Sep-1999  
 C:Accession: A46019; S25144  
 R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G  
 Genomics 15, 259-264, 1993  
 A>Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog  
 A:Reference number: A46019; MUID:93194170  
 A:Accession: A46019  
 A>Status: not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-2531 <DEL>  
 A:Cross-references: GB:211886; GB:S47228; NID:9288502; PIDN:CAA77941.1; PID:9288503  
 A>Note: sequence extracted from NCBI backbone (NCBI:P.127318)  
 R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenpan, R.J  
 submitted to the EMBL Data Library, April 1992  
 A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, sugg  
 A:Reference number: S25144  
 A:Accession: S25144  
 A:Molecule type: mRNA  
 A:Residues: 1551-2108, 'O', 2110-2114, 'ALP', 2118-2170 <PRA>  
 A:Cross-references: EMBL:211886  
 C:Genetics:  
 A:Gene: notch-1  
 A:Map position: 2  
 A>Note: proximal region of chromosome 2  
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol  
 F:106-138/Domain: EGF homology <EGF1>  
 F:144-175/Domain: EGF homology <EGF1>  
 F:222-254/Domain: EGF homology <EGF2>  
 F:261-292/Domain: EGF homology <EGF2>  
 F:339-370/Domain: EGF homology <EGF3>  
 F:416-449/Domain: EGF homology <EGF3>  
 F:456-487/Domain: EGF homology <EGF4>  
 F:494-525/Domain: EGF homology <EGF5>  
 F:532-563/Domain: EGF homology <EGF6>  
 F:607-638/Domain: EGF homology <EGF7>







Db 202 TCLAISCTLNGGCGHQCQVLTIVTHRCRCRPOYOLQEDGRRCVRSPCAEGNGGCMHI 261  
QY 203 CKTCSSGLTNRDCGCEHGVVL--DEGACVDVDECAAEPPCSAAOFCKNANGSY----- 254  
Db 262 CQELRLGLAH--CG-CHPQYOLADRKTEDEVCALGLAQ--AHGCLNTGSGFRVCYCHA 316  
QY 255 -----TCECDSSSCV--GCTGEGPGNCKECISGYAREHQ--CADV 291  
Db 317 GYELGADGROCRIEMELVNSCEANGSGCSHSTSTSPGLCTCPRGYELEDQKTCIDI 376  
QY 292 DECLAERTCYRKNECYNTPGSVVCVCPDGFEEEDAC 330  
Db 377 DDC--ANSPCQ--QACANTPGGYECSCFAGYRLNTDQC 411

RESULT 38  
A56136  
Jagged protein precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 28-Apr-1995 #sequence\_revision 28-Apr-1995 #text\_change 11-Jan-2000  
C:Accession: A56136  
R:Lindseil, C.E.; Shawber, C.J.; Boulter, J.; Weimaster, G.  
Cell 80, 909-917, 1995  
A:Title: Jagged: a mammalian ligand that activates Notch1.  
A:Reference number: A56136; MUID:95211842  
A:Accession: A56136  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1220 <LIN>  
A:Cross-references: GB:L38483  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
F:379-410/Domain: EGF homology <EGF1>  
F:492-523/Domain: EGF homology <EGF2>  
F:634-665/Domain: EGF homology <EGF2>

Query Match 11.2%; Score 224; DB 2; Length 1220;  
Best Local Similarity 22.8%; Pred. No. 5.5e-07;  
Matches 97; Conservative 42; Mismatches 101; Indels 186; Gaps 28;

QY 42 NQAVD-----TAKKNGGNTAMEEKTLSKYESSERLEILEGLCESS--DFECNQ 93  
Db 158 HSGMINSROWQTLKQNTG-----IAHFE-YQIRV-----TCDDHYGFGCKKF 200  
QY 94 LEAQBHLAAMWLKSEYPLF--EMPCVTKLKYCCSPGTGPPC--LACGGSGRPPSG 150  
Db 201 CRPD-----DFGHYACDQNGKNTCEGMGMPCCNKAIKQG-----CSP 241  
QY 151 NGHCSGDGSRGDSRCRHMGYOGPLCTDCM--DGYFSSLNNE----- 191  
Db 242 K-----HGSKLPGDCRQYQWQGLYCDKCIHPGCVATGCTNEPMQCLCTINMGQLCDK 296  
QY 192 -----THSTI-----TACDF-----SCK-TC 206  
Db 297 DLNYGTHQPLNKGTSNTGPRKYCCSPGYSGPNCIEAHACLSDPCHNRGSKETS 356  
QY 207 SGLNROGCEGVAV-----LDE-----GAC-----V 229  
Db 357 SGF---EC-ECSPGWTGTCSTNTIDDCSPNNCSHGCTQDLVNGFKVCYCPQWTKTQCL 412  
QY 230 DVDECAEPPCSAAOFCKNANGSYTC-----EECDSSSCVCGTC--EGPGNCKE--- 276  
Db 413 DANCEAK--PCVNARSKNLIASYCCLPGWNGQNDININDLGGCQNDASRDLYN 470  
QY 277 -----CISGYAREHQCADVDECLAERTCYRKNECYNTPGSVVCVCPDGF-----ET 326  
Db 471 GYRCICPGYAGDHCE--RDIDEC--ASNPL--NGHCCONEINRFQCLCPTGFSGNLCOLD 526  
QY 327 EDACVP 332  
Db 527 IDYCEP 532

RESULT 39  
A57172  
probable hormone receptor EMRI precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 11-Jan-2000  
C:Accession: A57172  
R:Baud, V.; Chissee, S.L.; Vlegas-Pequignot, E.; Dirlong, S.; N'Guyen, V.C.; Roe, B.A.  
Genomic 26, 334-344, 1995  
A:Title: EMRI, an unusual member in the family of hormone receptors with seven transmembrane domains  
A:Reference number: A57172; MUID:95324926  
A:Accession: A57172  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-886 <BAU>  
A:Cross-references: GB:X81479; NID:g784993; PIDN:CAA57232.1; PID:g784994  
C:Genetics:  
A:Gene: GDB:EMRI  
A:Cross-references: GDB:378349; OMIM:600493  
A:Map position: 19p13.3-19p13.3  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: transmembrane protein  
F:136-170/Domain: EGF homology <EGF>

Query Match 11.1%; Score 222.5; DB 2; Length 886;  
Best Local Similarity 27.9%; Pred. No. 5.3e-07;  
Matches 79; Conservative 20; Mismatches 95; Indels 89; Gaps 17;

QY 130 PGTYPDCLACQGSQRPSCNGHCSGDSRQDGSRCRHMGYOGPLCTDCM---DGYFS 186  
Db 74 PGVNRKIDIDEC--SQSPQCGPNSCK---NLGGRKCSCLDGFSPGTNDVPGAPGNS 129  
QY 187 SLRNEHSTICTACDESCCTCTGSLTNRDCG-----CEGVWVLEAGACVDV 232  
Db 130 -----CTDINE---C--LTSVRCPEHSDCVNSGYSQCVGFISRNSTCEDVN 174  
QY 223 ECAEPPCSAAOFCKNANGSYTC-----EECDSSC---V 264  
Db 175 EC-ADPRACPRHATCNNTVGNYSFCNPFESSGHLSCQGLAKSCEDIIDCTEMCPINS 233  
QY 265 GCTGEGPGN--CKECISGYAREHQ-----CADVDECSLAERTCYRKNECYNTPG 313  
Db 224 TCTNT--PGSYPC--TCHPFAFSSQGLNFTDQGVRECRDIDECROPSTC--GPNISICTNALG 290  
QY 314 SYVCVCPDGFEEEDACVPAEAETEGESPQT---LPSRDL 353  
Db 291 SYSCGCTIVGFH-----PNEGSGKDGNFSCQRYLFXCKEDV 326

RESULT 40  
KXBOS  
Plasma protein S precursor - bovine  
N:Alternate names: vitamin K-dependent protein S  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 08-Aug-1987 #sequence\_revision 26-Jan-1996 #text\_change 16-Jul-1999  
C:Accession: A24759  
R:Dahlback, B.; Lundvall, A.; Stenflo, J.  
Proc. Natl. Acad. Sci. U.S.A. 83, 4199-4203, 1986  
A:Title: Primary structure of bovine vitamin K-dependent protein S.  
A:Reference number: A24759; MUID:86233400  
A:Accession: A24759  
A:Molecule type: mRNA  
A:Residues: 1-675 <DNA>  
A:Cross-references: GB:M13044; NID:g163697; PIDN:AAA30757.1; PID:g163698  
A:Note: Parts of this sequence, including the amino end of the mature protein, were determined from the complementary DNA sequence of the cDNA clone pSVL-1.  
J. Biol. Chem. 261, 5111-5115, 1986  
A:Title: Localization of thrombin cleavage sites in the amino-terminal region of bovine plasma protein S.  
A:Reference number: A23888; MUID:86168236  
A:Accession: A23888  
A:Molecule type: protein  
A:Residues: 42-141 <DA2>  
C:Complex: In plasma forms a complex with C4b binding protein

C:Function:  
 A:Description: a cofactor for activated protein C (EC 3.4.21.69); thrombin cleavage dest  
 C:Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat hom  
 C:Keywords: beta-hydroxyaspartate; beta-hydroxyaspartic acid; blood coagulation; carbox  
 F:1-24/Domains: signal sequence #status predicted <SIG>  
 F:25-41/Domains: propeptide #status predicted <PRO>  
 F:26-85/Domains: Gla domain homology <Gla>  
 F:42-67/Products: plasma protein S #status experimental <MAT>  
 F:121-154/Domains: EGF homology <EG1>  
 F:161-199/Domains: EGF homology <EG2>  
 F:205-241/Domains: EGF homology <EG3>  
 F:247-282/Domains: EGF homology <EG4>  
 F:315-666/Domains: sex hormone-binding globulin homology <SHB>  
 F:325-478/Domains: laminin G repeat homology <LGR>  
 F:47-48-55-57-60-61-66-67-70-73-77/Modified site: gamma-carboxyglutamic acid (Glu) #stat  
 F:58-63-88-113-121-134-126-143-145-154-161-175-171-184-186-199-205-217-212-226-241,2  
 F:93-94/Cleavage site: Arg-Ala (thrombin) #status experimental  
 F:111-112/Cleavage site: Arg-Ser (thrombin) #status experimental  
 F:136/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental  
 F:177-219-258/Modified site: erythro-beta-hydroxyaspartic acid (Asn) #status experimental  
 F:288-567-449-475-638-665/Disulfide bonds: #status experimental  
 F:499/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:509/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.1%; Score 222; DB 1; Length 675;  
 Best Local Similarity 26.4%; Pred. No. 4-6e-07;  
 Matches 83; Conservative 29; Mismatches 98; Indels 104; Gaps 20;

OY 57 NTAMEKTSKYSSRIRLEILEGICSSD-----FECNOMLEAOEHEHAWMLQ----- 108  
 DB 43 NTLLEETKKGNLER-----ECIEELCNKEAREIEENN---PELEYEPKYLGLGIGSF 92  
 OY 109 -----KSEYDLEFEMFCVKTLKVCSPGTYGPDCLACQGSORPCSGNGHCS-G 156  
 DB 93 RAGLFRAALSTNAYPDLNS--CVNAISPOCN-----LPENNDGEMTCK 135  
 OY 157 DGSRGDGSRCRCHMGYQPLC---TDCMDGYFSSLRNETHSICTACDESKCTSGLTNR 212  
 DB 136 DG--QATFTICIKSGWQGEKCESDINECKDPV-----NINCGCSQICENTPG--SY 182  
 OY 213 DCGCEGVNVL--DEGACVDVDECAAEPPPCSAOFCCKNANGSYTEDECDSSVCGTGBG 270  
 DB 183 HC-SCNNGEVMNSKKDKCDVDECVLKPSICGA-VCKNIPGPFEC----- 226  
 OY 271 PGNCKECIGYARE--HGOCADVDECSLAETCVRKNECYNTPGSYVCV--PDGFEET 326  
 DB 227 -----ECARGYKKNPYKSKCDDVDEC--AENICA---QLCVNTPGISTCYCDGKRGFKLA 276  
 OY 327 ED-----ACVP 332  
 DB 277 ODOKSCEAVPCLP 290

RESULT 41  
 T20125  
 Hypothetical protein C50H2.3a - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T20125  
 R:McMurray, A.  
 submitted to the EMBL Data Library, June 1996  
 A:Reference number: Z19226  
 A:Accession: T20125  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-838 <M11>  
 A:Cross-references: EMBL:Z73971; PIDN:CAA98251.1; GSPDB:GN00023; CESP:C50H2.3a  
 A:Experimental source: clone C50H2  
 C:Genetics:  
 A:Gene: CESP:C50H2.3a  
 A:Map position: 5  
 A:Intons: 31/1; 98/1; 162/1; 199/1; 244/1; 279/1; 323/1; 354/2; 391/1; 408/1; 458/1; 48

Query Match 11.1%; Score 222; DB 2; Length 838;  
 Best Local Similarity 27.2%; Pred. No. 5-5e-07;  
 Matches 75; Conservative 25; Mismatches 92; Indels 84; Gaps 17;

OY 117 EMEFCVTLKVCSPGTYGPDCLACQ-----GSQRPC-----SG 150  
 DB 51 QMFWDQVSE--KEFHYG-----GCMQTKNRFSSKQCKRYKMFNPAVPDLCLLDAD 105  
 OY 151 NCHSGDGS-----ROQDSCRCRCHMGYQPLCTDC--MDGYFSSL-----RNFT 192  
 DB 106 QGHGDERNGHMYFPNDSGECEKFFY-----GCGNDKFFSLHMKCRVCGERLSP 159  
 OY 193 HSICFACD---ESCKTCSGLTNRDCECEVNWLDG--GACVDVDECAEPPPCSAOFCCK 248  
 DB 160 QIACHDCLRTSTCKSNSNF--NTYC--ECRSYGEKNYGCIDIDECRGKAVCDRANCV 217  
 OY 249 NANGSYTCECDSSCVG-----CTGEGPG-----NCKECI-----SGYAREHQ 287  
 DB 218 NEIGSYKC--ECMASYRGDKHCTYVGLGRSSIDCKDKMSHATCMNGVCQCKEGYEGDGFN 276  
 OY 288 CADVDECSLAETCVRKNECYNTPGSYVCPPDPF 323  
 DB 277 CTDVNCELRPRPWC--NKNAECTINRGSTICTCLEGT 311

RESULT 42  
 A38261  
 Masking protein precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 21-Jan-2000  
 C:Accession: A38261  
 R:Tsuji, T.; Okada, F.; Yamaguchi, K.; Nakamura, T.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 8835-8839, 1990  
 A:Title: Molecular cloning of the large subunit of transforming growth factor  
 A:Reference number: A38261; MUID:91062373  
 A:Accession: A38261  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1712 <TSU>  
 A:Cross-references: GB:M55431; NID:9207285; PIDN:AAA42235.1; PID:9207286  
 C:Superfamily: unassigned EGF-related proteins; EGF homology; LDL receptor ligand-bl  
 F:911-947/Domains: EGF homology <EGF>

Query Match 11.1%; Score 222; DB 2; Length 1712;  
 Best Local Similarity 31.1%; Pred. No. 9-6e-07;  
 Matches 68; Conservative 19; Mismatches 62; Indels 70; Gaps 15;

OY 150 GNGHCSGDSRGDSCRCRCHMGYQPLCTDCMDGYFSSLRNETHSICTACDESKCTSGL 209  
 DB 877 GAGHCINLPVRY--TCIYEGYK-----FS-----EQQRKCIDIE--CAQAQL 917  
 OY 210 TNRDCECE-----VGWVL--DEGA--CYDVDECAEPPPCSAOFCCKNANGSYTC 256  
 DB 918 CSQ--GRCENTEGSLICLPAGFIASEBGSNCIDVDEC--LRDPVCDGR--CINTAGAFRC 973  
 OY 257 EEDDS-----SCVGTGCEBPGN--CKECSGYAREHGQA 289  
 DB 974 EYCDSGYRMSRGRCHDIDECILPTSCPBEQCY-----NSPFGSYQVCTEGFRGNNGQL 1029  
 OY 290 DVDECSLAETCVRKNECYNTPGSYVCPPDGFETED 328  
 DB 1030 DVDEC-LQPKVCA--NGSCTNIEGSMCSCHNGYSPTPD 1065

RESULT 43  
 A49175  
 Notch B protein - mouse (fragment)  
 N:Alternate names: Notch homolog  
 C:Species: Mus musculus (house mouse)  
 C:Date: 21-Jan-1994 #sequence\_revision 05-Jan-1996 #text\_change 20-Sep-1999

C:Accession: A49175; PH1570; S32113  
R:Lardelli, M.; Lendahl, U.  
Exp. Cell Res. 204, 364-372, 1993  
A:Title: Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of  
A:Reference number: A49175; MUID:93178563  
A:Accession: A49175  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1203 <LAR>  
A:Cross-references: EMBL:668279; NID:9287989; PIDN:CAA48340.1; PID:9287990  
A:Experimental source: embryo  
A:Note: sequence extracted from NCBI backbone (NCBI:126158)  
C:Comment: This protein has many EGF repeats and 11n-12/Notch repeats.  
C:Superfamily: This protein is one of the neurogenic proteins controlling the decision between  
F:143-174/Domain: EGF homology <EGF1>  
F:148-513/Domain: EGF homology <EGF1>  
F:560-591/Domain: EGF homology <EGF1>  
F:674-705/Domain: EGF homology <EGF2>  
F:712-743/Domain: EGF homology <EGF3>  
F:836-867/Domain: EGF homology <EGF3>

Query Match 11.0%; Score 221; DB 2; Length 1203;  
Best Local Similarity 22.4%; Pred. No. 8.4e-07;  
Matches 91; Conservative 44; Mismatches 113; Indels 158; Gaps 24;

OY 28 PTPCH-RCRLVDKFNQMGVDTAKKNGGNTAMEETLSKYES---SEIRLEITLGL 82  
DB 259 PDPCHHGCGQDIDSY-----TTCINPGYGALICSDIDBCYSPCLNDGICDILVNG- 311  
OY 83 CESSDFECNQMLEAEHELEAMWLQKSEYDLEFEMFCVKTLKVCSPGTGPPDC-LACQ 141  
DB 312 -----YQCN-----CQGTSLNCEINFD 330  
OY 142 GGSORPCSGNCHSGDGRQDSCRCMHGYOGPLC-----TDCMDGYFSSLNET 192  
DB 331 DCASIPFCM-HGVCDVGINRY---SCVCSPTGQRNCIDIDECASNPCRKG--ATCINDV 384  
OY 193 HSICTAODE-----SCKT-----CS-GLTNRQCGEEVGVW-----LDGGA 227  
DB 385 NGFRICCEGPHHPSQVSYVNECLSNPCIGHNCTGGLGYKC-LCDAGVGVNCEVDNE 443  
OY 228 C-----VDVDECAAPPCSAAPFCCKNANGSYT 255  
DB 444 CLSNPCQNGCTCNLYNVRCTCKKGFKGYNCQVNIIDECASN--PCLNMGTCFDVYSGT 501  
OY 256 C-----EECDSSCVGCTG---EGPNCRE-----CISGYAREHGQC-ADVDSCS 295  
DB 502 CHCMLPYTGKNCQVTLACSPNCPENAVKCAPNFESFSLCAPMGCKRCTVDVDC- 560  
OY 296 LAERTCVAKKNCYTPPSYVCVCGDF-----EETEDACVPAPE 335  
DB 561 -ISKPCM-NGVCHNTQGSYVCECPGFSGMDCEDINDCLANPCQ 604

## RESULT 44

## MMHUND

nidogen precursor - human  
N:Alternate names: entactin  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 20-Oct-2000  
C:Accession: A33322; A32437; A61367  
R:Nagayoshi, T.; Sanborn, D.; Hickok, N.J.; Olsen, D.R.; Fazio, M.J.; Chu, M.L.; Knowlton  
DNA 8, 581-594, 1989  
A:Title: Human nidogen: complete amino acid sequence and structural domains deduced from  
A:Reference number: A33322; MUID:90091745  
A:Accession: A33322  
A:Molecule type: mRNA  
A:Residues: 1-1247 <NAG>  
A:Cross-references: EMBL:M30269  
R:Olsen, D.R.; Nagayoshi, T.; Fazio, M.; Mattei, M.G.; Passage, E.; Weil, D.; Timpl, R.;  
Am. J. Hum. Genet. 44, 876-885, 1989

A:Title: Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to  
A:Reference number: A32437; MUID:89270475  
A:Accession: A32437

A:Molecule type: mRNA  
A:Residues: 667-1247 <OLIS>

A:Cross-references: EMBL:M27445; NID:9602466; PIDN:AAA57261.1; PID:9602467

A:Note: the authors translated the codon AAG for residue 966 as Cys

R:Fazio, M.J.; O'Leary, J.; Kaehner, V.M.; Chen, Y.Q.; Salter, B.; Uitto, J.

J. Invest. Dermatol. 97, 281-285, 1991  
A:Title: Human nidogen gene: structural and functional characterization of the 5'-fla

A:Reference number: A61367; MUID:91302882

A:Accession: A61367  
A:Molecule type: DNA

C:Comment: This protein is a basement membrane glycoprotein that forms a complex with

C:Genetics:  
A:Gene: GDB:NID

A:Cross-references: GDB:120236; OMIM:131390  
A:Map position: 1q43-1q43

C:Superfamily: nidogen: EGF homology; LDL receptor WYTD-containing repeat homology; t  
C:Keywords: basement membrane; beta-hydroxyasparagine; calcium binding; cell binding;

protein  
F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1247/Product: nidogen #status predicted <MAT>

F:390-425/Domain: EGF homology <EG1>

F:672-708/Domain: EGF homology <EG2>

F:702-704/Region: cell attachment (R-G-D) motif

F:714-750/Domain: EGF homology <EG3>

F:762-800/Domain: EGF homology <EG4>

F:806-839/Domain: EGF homology <EG5>

F:849-919/Domain: thyroglobulin type I repeat homology <THY1>

F:990-1032/Domain: LDL receptor WYTD-containing repeat homology <WY1>

F:1033-1075/Domain: LDL receptor WYTD-containing repeat homology <WY2>

F:1076-1120/Domain: LDL receptor WYTD-containing repeat homology <WY3>

F:1121-1160/Domain: LDL receptor WYTD-containing repeat homology <WY4>

F:1161-1197/Domain: LDL receptor WYTD-containing repeat homology <WY5>

F:1212-1243/Domain: EGF homology <EG6>

F:1289-1296/Modified site: sulfate (Tyr) (covalent) #status predicted

F:729-819/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:755/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino of laminin) #status pr

F:1137/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.0%; Score 221; DB 1; Length 1247;  
Best Local Similarity 27.1%; Pred. No. 8.7e-07;  
Matches 54; Conservative 28; Mismatches 75; Indels 42; Gaps 7;

OY 174 GPLCTCDMDGYFSSLRNETHSICITACDESKTSGLTNRDGECEVGLDEGACVDVE 233  
DB 658 GPV-----REGSPDALQNCYIGTHGCDINMACRPFRQFCEGIGRGGORTCYDIDE 713

OY 234 CAEPPPCSAAPFCNANGSYTCEEC-----DSSCVGCTGEGP-----GNCK-- 275  
DB 714 CSEQPSVCGSHITCNHGHGFRFC-ECVEGYQFSGDGTCAVAVDQRPINVCETGLHNCIP 772

OY 276 -----ECISGYAREHGQADVDECSLAETCKRKKNKNCYNPPGYSVYCP 320  
DB 773 ORAQCIYTGSSYTCSCLPFGSGDQACADVDEC---QPSRCHPAFCYNTFGSFTCKCK 829

OY 321 DGEETEDACVPAPEAAT 339  
DB 830 PGYQGDGFRVC-PGEVEKT 847

RESULT 45

A40043

notch protein homolog TAN-1 precursor - human

C:Species: Homo sapiens (man)

C:Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 13-Aug-1999

C:Accession: A40043

R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Skl

Cell 66, 649-661, 1991

A:Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromoso

A:Reference number: A40043; MUID:91347367  
 A:Accession: A40043  
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A:Molecule type: mRNA  
 A:Residues: 1-2555 <ELL>  
 A:Cross-references: GB:M73980  
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology  
 F:261-292/Domain: EGF homology <EGF1>  
 F:494-525/Domain: EGF homology <EGF1>  
 F:987-1018/Domain: EGF homology <EGF2>  
 F:1149-1180/Domain: EGF homology <EGF3>  
 F:1187-1218/Domain: EGF homology <EGF3>  
 F:1233-1264/Domain: EGF homology <EGF3>  
 F:1927-1959/Domain: ankyrin repeat homology <AN1>  
 F:1960-1992/Domain: ankyrin repeat homology <AN2>  
 F:1994-2026/Domain: ankyrin repeat homology <AN3>  
 F:2027-2059/Domain: ankyrin repeat homology <AN4>  
 F:2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 11.0%; Score 221; DB 2; Length 2555;  
 Best Local Similarity 22.4%; Pred. No. 1.5e-06;

Matches 79; Conservative 20; Mismatches 105; Indels 148; Gaps 16;

QY 125 KVCSPFGYGPDLACQGSQRPSCSGNGHCSGDSRQDGSRCRHMGYQPLCTD-----179  
 DB 126 KCRCPGWSKSCQADPCASNPANGQCL---PEASYICHPSPFHGPTCRQDVNEC 182  
 QY 180 -----CMGYSFLNETHS---ICTA-----CDESCSTCS-----207  
 DB 183 GQKPRLCRHH--GTCHNEVGYRCVCRATHHTGPNCEPVPSPSPQNGTCRPTGDTV 240  
 QY 208 -----GLTRDCE-----CEVGWVLDGACV-----D 230  
 DB 241 HECACLPFGTQGCENIDDCPGNNCKNG-----GACVDSVNTYNGPCRPETWGTQCTED 295  
 QY 231 VDECAAPPPCSAAOFCKNANGSYTC-----EBCDSCVCGCTGEG--PENCKE---276  
 DB 296 VDECAAPPPCSAAOFCKNANGSYTC-----EBCDSCVCGCTGEG--PENCKE---276  
 QY 277 -----CISGVAREHQG-----ADVDE 293  
 DB 356 SFCECPHGTGLCHLNDACISNPNESNCTNPYNGKALICTCSGYTGPAQSDVD 415  
 QY 294 CSIAERTCYARKNENCTPGSYVCPDGF-----EETEDACVP-PAEAET 339  
 DB 416 CSIGANPCEHAGK-CINTLGSFECQCLQGYTGPRCEIDVNECVSNPCQNDAT 466

RESULT 46

S18188

notch protein homolog - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999

C:Accession: S18188

A:Title: A homolog of Drosophila Notch expressed during mammalian development.

A:Reference number: S18188; MUID:92111383

A:Accession: S18188

A:Molecule type: mRNA

A:Residues: 1-2531 <WEI>

A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

F:987-1018/Domain: EGF homology <EGF1>

F:1025-1056/Domain: EGF homology <EGF2>

F:1233-1264/Domain: EGF homology <EGF2>

F:1917-1949/Domain: ankyrin repeat homology <AN1>

F:1950-1982/Domain: ankyrin repeat homology <AN2>

F:1984-2016/Domain: ankyrin repeat homology <AN3>

F:2017-2049/Domain: ankyrin repeat homology <AN4>

F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 11.0%; Score 220; DB 2; Length 2531;  
 Best Local Similarity 23.6%; Pred. No. 1.7e-06;  
 Matches 76; Conservative 23; Mismatches 97; Indels 126; Gaps 17;

QY 126 VC-CSFGTYGPDL-----ACQGSQRP--CSGNGHCSGDSRQDGSRCRHMGYQPLC-177  
 DB 163 ICQCPGFGHPTCRQDVNEC---SQNPGLCRHGTCHNE---IGSYRCACRATHGPHCE 216  
 QY 178 -----TDGMDGYFSSLRNETHSICTA-----CDESCSTCSGLTRDCECEVM 221  
 DB 217 LRPVPCSPPCQNGTCRPTGDTTHCACLPFAGNCEENVDGCG-----NNCKNG- 269  
 QY 222 VDEGACV-----DVDECAAPPPCSAAOFCKNANGSYTC-----256  
 DB 270 -----GACVGVNNTYNGRCPEPTWGTQCTEDVDECAQMPNACQAGTCHNSHGNCVCVN 325  
 QY 257 -----EBCDSCVCGCTGEG--PENCKE-----CISGVAREHG 286  
 DB 326 GWTGEDCSNIDDCASAGFCOGATCHDRVASFCECPHRTGLCHLNDACISNPNESGS 385  
 QY 287 QC-----ADVDECSIAEKTGYARKNENCTPGSYVCPDGF 323  
 DB 386 NCDTNPYNGKALICTCPRGYTGPAQSDVDDECAIGANPCEHAGK-CINTLGSFECQCLQGY 444  
 QY 324 -----EETEDACVP-PAEAET 339  
 DB 445 TGPRCEIDVNECVSNPCQNDAT 466

RESULT 47

T08618

intrinsic factor-B12 receptor CUBILIN precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: T08618

R:Moestrup, S.K.; Kozyraki, R.; Kristiansen, M.; Kayser, J.H.; Rasmussen, H.H.; Braul

J. Biol. Chem. 273, 5235-5242, 1998

A:Title: The intrinsic factor-B12 receptor and target of teratogenic antibody

A:Reference number: 216459; MUID:98148073

A:Accession: T08618

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3623 <MOD>

A:Cross-references: EMBL:AF022247; NID:g3834379; PIDN:AAC71661.1; PID:g3834380

C:Genetics:

A:Gene: CUBILIN

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membra

F:1-20/Domain: signal sequence #status predicted <Sig>

F:21-3623/Product: Intrinsic factor-B12 receptor CUBILIN #status predicted <Mat>

F:133-164/Domain: EGF homology <EGF1>

F:436-467/Domain: EGF homology <EGF2>

A:Accession: T08618

A:Molecule type: mRNA

A:Residues: 1-3623 <MOD>

A:Cross-references: EMBL:AF022247; NID:g3834379; PIDN:AAC71661.1; PID:g3834380

C:Genetics:

A:Gene: CUBILIN

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membra

F:1-20/Domain: signal sequence #status predicted <Sig>

F:21-3623/Product: Intrinsic factor-B12 receptor CUBILIN #status predicted <Mat>

F:133-164/Domain: EGF homology <EGF1>

F:436-467/Domain: EGF homology <EGF2>

A:Accession: T08618

A:Molecule type: mRNA

A:Residues: 1-3623 <MOD>

A:Cross-references: EMBL:AF022247; NID:g3834379; PIDN:AAC71661.1; PID:g3834380

C:Genetics:

A:Gene: CUBILIN

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membra

Db 302 DINECEINNGSGQAPLPLCLTPGSPSCNCPAGSGDGRVCTPVDICSIHNGCHPEA 361  
Oy 339 TEGESP 344  
Db 362 TCSSSP 367

## RESULT 48

cell-fate determining gene Notch2 protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 13-Aug-1999  
C:Accession: A49128  
R:Meinaster, G.; Roberts, V.J.; Lemke, G.  
Development 116, 931-941, 1992  
A:Title: Notch2: a second mammalian Notch gene.  
A:Reference number: A49128; MUID:93202015  
A:Accession: A49128  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-2471 <ME1>  
A:Experimental source: Schwann cell  
A:Note: Sequence extracted from NCBI backbone (NCBIP.127811)  
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology  
F:264-295/Domain: EGF homology <EGX1>  
F:799-830/Domain: EGF homology <EGF1>  
F:877-908/Domain: EGF homology <EGX2>  
F:1029-1060/Domain: EGF homology <EGF>  
F:1067-1098/Domain: EGF homology <EGF3>  
F:1153-1184/Domain: EGF homology <EGF3>  
F:1191-1222/Domain: EGF homology <EGF4>  
F:1876-1908/Domain: ankyrin repeat homology <AN1>  
F:1909-1941/Domain: ankyrin repeat homology <AN2>  
F:1943-1975/Domain: ankyrin repeat homology <AN3>  
F:1976-2008/Domain: ankyrin repeat homology <AN4>  
F:2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 10.9%; Score 218; DB 2; Length 2471;  
Best Local Similarity 22.6%; Pred. No. 2.3e-06;

Matches 92; Conservative 43; Mismatches 112; Indels 160; Gaps 24;

Oy 28 PTPCH--RCRGLVDKFNQGNVDTAKKNGGNTAMEEKTLSKYS--SEIRLEILEGL 82  
Db 576 PDPCHHGCGGIDSY-----TCICNNGYMGALICSDQIDCYSPLCNDRCIDLVNG- 628  
Oy 83 CESSDFECNQMLEAOEHLLEAMWLQKSEYDLEFEMFCVTKLVCCSPGYGPPDC-LACQ 141  
Db 629 ----YQCN-----CQPGTSLNCEINFD 647  
Oy 142 GGSQRPSGNGHSGDGRGDSRCRMGYQGPLCTDCMDGYFS-----SLRNETHS 194  
Db 648 DCASNPLI-HGACVDGIRY--SCVCSPGFTGRCNIDIDECASNPCRKATCLINDVNG 703  
Oy 195 ICTACDE-----SCKT-----CS-GLTNRDCECEVGNV-----LDEGAC- 228  
Db 704 FRCHCPREPHRPSGYOYNECLSPICINGTGSLGYKC-LDAGWVGINCEVDKNCL 762  
Oy 229 -----VDYDECAEPSPCSAOPFCNNANGSYTC- 256  
Db 763 SNPONGGTCNNLVNGYRCTCKKGFKGYNCOVINIDECASN--PCINOGTCLDDVSGYTC 820  
Oy 257 -----EECDSSCVGCTG--EGPGNCKE-----CISGVARHQG--CADVDEC 294  
Db 821 CMLEPYTGKNCOTVLAPSPSPNCENANVCKEAPNFESFTCLDAPG--OGRCTYVDDEC 877  
Oy 295 SLAEKTCVRKNENYNTPGSYVVCVCPDGF-----EETEDACVPPAE 335  
Db 878 --VSKPCM--NNGICHTNNGSGYMCBPPGFSMDCEBDINDCLANPCQ 921

RESULT 49  
T16860

hypothetical protein T13C2.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Aug-2000  
C:Accession: T16860

R:Du, Z.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid T13C2.  
A:Reference number: 218591

A:Accession: T16860

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1357 <DU2>

A:Cross-references: EMBL:U00030; NID:g1055164; PID:g1055166; PIDN:AAA81134.1; CESP:T1

A:Gene: CESP.T13C2.4

A:Introns: 38/2; 66/2; 115/2; 158/3; 233/1; 287/3; 325/3; 350/2; 432/2; 502/2; 525/1;

C:Superfamily: LDL receptor ligand-binding repeat homology

F:526-561/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:565-609/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:614-653/Domain: LDL receptor ligand-binding repeat homology <LDL3>

F:660-699/Domain: LDL receptor ligand-binding repeat homology <LDL4>

F:706-741/Domain: LDL receptor ligand-binding repeat homology <LDL5>

F:748-784/Domain: LDL receptor ligand-binding repeat homology <LDL6>

F:791-836/Domain: LDL receptor ligand-binding repeat homology <LDL7>

Query Match 10.8%; Score 217.5; DB 2; Length 1357;  
Best Local Similarity 22.7%; Pred. No. 1.5e-06;

Matches 121; Conservative 43; Mismatches 141; Indels 229; Gaps 32;

Oy 1 MRLRRRAA--LGLLP-----LTLPLPAPAEAKKPT-----PCHR---CRGLVD 39  
Db 441 VNLNRKSHGFLPEIYVTEERKLCULLPPRRRRSVTQNTPIHKSVCQQLVYL 500  
Oy 40 KFNQGNVDTAKKNGGNTAMEEKTLSKYSSEIRLEILEGLCESSDFECNQMLEAO-- 97  
Db 501 SSRIVREHVRVYPRFKNI---KTLIR-ECDATNSFOCDGRCIPMSWRCDGIDICONE 555  
Oy 98 -----EHLLEAMWLQKSEYDLEFEMFCVTKLVCC- 127  
Db 556 EDEKNCPRVGAEHKG--EVKSARSLSERFCIPKRWCVDEDFEDCDKSDPECKNV 612  
Oy 128 -----CS--PGTYG--PDCLACOGSGSORPSGNG--HSGDGRGDSRCRHM 170  
Db 613 SCQKQKQFCBELSDGYSICITETWVCDG--QRDCT-NKQDKQNTSKTSKCPDNNFQCSN 669  
Oy 171 G--YQGPLC--TDCMDGYFSSLR---NET-----TCS-GLTNRDCE- 192  
Db 670 GNCIFKNWVCGDEDCSDGDELTLAPSNCRVYNOCPGEMMKKSGSECIPSRMRDAE 729  
Oy 193 -----HS--ICTACDECK-----TCS-GLTNRDCE- 215  
Db 730 VDKDHSDEKNCVTAIQHTCKLAEEFACKASHNCINKAFVCDGELDCSDGSDDECDADVRT 789  
Oy 216 ECEVGS-----VWLDGACVVDV----- 232  
Db 790 ECKSGERTCPASYAYGAESEHVCIPASSMKNGBEPPDGDDEKCMATAPYCOQKTE 849  
Oy 233 -ECAAP-----PCSAQF--CKNANGSYTC-----EECDSSCVGCT-----GEG 270  
Db 850 YECSTPLOCITKMSKLCASAOFGDGNMS--VCSOKKTIEMCKPSSSEGCVRPSPVRGNN 908  
Oy 271 PGNCKECSYGAREHGOCADVDGCSLAKEKTCVRKNENYNTPGSYVVCVCPDGF 324  
Db 909 VCHCKD--GYKLENGQCIDINECEIA-GVC--DQICLNPISGYRCACHAGYQ 955

RESULT 50  
T23433  
hypothetical protein K08C7.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T23433



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 10:58:58 ; Search time 13.17 Seconds  
(without alignments)  
654.688 Million cell updates/sec

Title: US-09-905-075-2

Perfect score: 2005

Sequence: 1 MRLPRRAAGLPLLLLP.....APEATEGESPTQLPSREDL 353

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*

2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*

3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*

4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*

5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*

6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	235.5	12.7	1253	US-08-479-722B-4	Sequence 4, Appl1
2	234.5	12.7	652	US-08-751-305-2	Sequence 2, Appl1
3	250	12.5	1833	US-08-479-722B-2	Sequence 2, Appl1
4	250	12.5	1833	PCT-US95-02251-18	Sequence 18, Appl1
5	246.5	12.3	1251	US-08-199-780-3	Sequence 3, Appl1
6	246.5	12.3	1252	US-08-316-650-3	Sequence 3, Appl1
7	240.5	12.0	1394	517197-30	Patent No. 517197
8	240.5	12.0	676	US-08-282-141-4	Sequence 4, Appl1
9	239	11.9	676	US-08-435-434-3	Sequence 3, Appl1
10	239	11.9	676	US-08-435-436-3	Sequence 3, Appl1
11	239	11.9	676	US-08-438-863-3	Sequence 3, Appl1
12	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
13	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
14	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
15	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
16	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
17	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
18	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
19	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
20	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
21	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
22	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
23	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
24	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
25	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
26	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1
27	239	11.9	676	US-08-438-864-3	Sequence 3, Appl1

## ALIGNMENTS

28	227.5	11.3	635	1	US-08-436-804-2	Sequence 2, Appl1
29	227.5	11.3	635	1	US-08-267-387-2	Sequence 2, Appl1
30	224	11.2	652	6	5258288-4	Patent No. 5258288
31	224	11.2	1219	4	US-08-882-046-5	Sequence 5, Appl1
32	223.5	11.1	288	1	US-08-368-852-15	Sequence 15, Appl1
33	223	11.1	2523	1	US-08-185-432-18	Sequence 18, Appl1
34	222.5	11.1	886	3	US-09-110-116-3	Sequence 3, Appl1
35	222	11.1	2556	1	US-08-185-432-17	Sequence 17, Appl1
36	222	11.1	2556	1	US-08-083-590A-20	Sequence 20, Appl1
37	222	11.1	2556	3	US-08-532-384-20	Sequence 20, Appl1
38	220.5	11.0	1010	4	US-08-882-046-7	Sequence 7, Appl1
39	220.5	11.0	1036	4	US-09-068-740A-7	Sequence 7, Appl1
40	220.5	11.0	1187	4	US-09-068-740A-7	Sequence 7, Appl1
41	220.5	11.0	1218	2	US-08-400-159-6	Sequence 6, Appl1
42	220.5	11.0	1218	3	US-08-611-729A-6	Sequence 6, Appl1
43	220.5	11.0	1218	4	US-08-882-046-2	Sequence 2, Appl1
44	220.5	11.0	1218	4	US-09-214-278-7	Sequence 7, Appl1
45	220.5	11.0	1218	4	US-09-068-740A-11	Sequence 11, Appl1
46	218.5	10.9	675	6	5258288-1	Patent No. 5258288
47	216.5	10.8	2703	1	US-08-185-432-19	Sequence 19, Appl1
48	215.5	10.7	2471	1	US-08-083-590A-19	Sequence 19, Appl1
49	215.5	10.7	2471	1	US-08-532-384-19	Sequence 19, Appl1
50	215.5	10.7	2471	3	US-08-532-384-19	Sequence 19, Appl1
51	213	10.6	673	1	US-08-282-141-3	Sequence 3, Appl1
52	213	10.6	673	1	US-08-435-434-1	Sequence 1, Appl1
53	213	10.6	673	1	US-08-435-436-1	Sequence 1, Appl1
54	213	10.6	673	2	US-08-438-863-1	Sequence 1, Appl1
55	213	10.6	673	2	US-08-438-864-1	Sequence 1, Appl1
56	213	10.6	673	2	US-08-438-864-1	Sequence 1, Appl1
57	213	10.6	673	4	US-08-628-747-1	Sequence 1, Appl1
58	213	10.6	673	4	US-08-402-253-1	Sequence 1, Appl1
59	213	10.6	673	4	US-08-443-866B-1	Sequence 1, Appl1
60	211.5	10.5	810	2	US-08-820-170A-34	Sequence 34, Appl1
61	211.5	10.5	810	2	US-09-055-699-34	Sequence 34, Appl1
62	211.5	10.5	810	4	US-09-273-565-34	Sequence 34, Appl1
63	211.5	10.5	810	4	US-09-565-538-34	Sequence 34, Appl1
64	209.5	10.4	448	2	US-08-884-072-1	Sequence 1, Appl1
65	209.5	10.4	448	4	US-09-212-168-1	Sequence 1, Appl1

RESULT 1  
US-08-479-722B-4  
Patent No. 6074840  
GENERAL INFORMATION:  
APPLICANT: Bonadio, Jeffrey  
APPLICANT: Yin, Mushan  
TITLE OF INVENTION: LATENT TGF ( BINDING PROTEIN (LTBP)  
TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Williams, Morgan & Amerson  
STREET: 7676 Hillmont, Suite 250  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77040  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479, 722B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US95/02251  
FILING DATE: 21-FEB-1995  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/316,650  
 FILING DATE: 30-SEP-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/199,780  
 FILING DATE: 18-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fussey, Shelley P.M.  
 REGISTRATION NUMBER: 39,458  
 REFERENCE/DOCKET NUMBER: 4100.000500/FUS  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (713) 934-7000  
 TELEFAX: (713) 934-7011  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1253 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-479-722B-4

Query Match 12.7%; Score 255.5; DB 3; Length 1253;  
 Best Local Similarity 24.0%; Pred. No. 3e-12;  
 Matches 102; Conservative 40; Mismatches 108; Indels 175; Gaps 25;

QY 20 PABEAAKPTPCRHCRGLVDKFNQGMVDTAKKNFGGNTAMEEKLTSKYSSEIRLEIL 79  
 DB 518 PYBELIRSP-----PTFHRFLPDLPP-----SRSAVEIAPQVETDECRINQNI 564  
 QY 80 --BGLC--ESSDF--ECNQMLEAQEHLFAMWQLKSEYDLEFEMCVKTLKCCSPGTY 133  
 DB 565 CGHGCVPGSPSDYSCHCMNGYRSHPOH-----YCVDVNE--CEAEPG 605  
 QY 134 GPDLCAQ--GGSORPCSGNGHS-----GDGSR-----GDG----- 164  
 DB 606 GPKGKICMTNGSY-----NCHCNRGYRLHVGAGRSQVDLNCARPHLCGGDGFPCINFP 660  
 QY 165 ---SCRHMGY---QGPICTD-----CMDGYFSSLNETHSI--CTACDESKTC 206  
 DB 661 GHYKCMCYPGYRLKASRPICEDIDECRDPSTCPDG--KCEMKPGSFKICACQPGYRSQ 717  
 QY 207 SGLTNDCCGEFVGMWDEACVDECAAEPPPCSAQAQCKNANGSYTC----- 256  
 DB 718 GG-----GACRDVNEC--SEGTPCSPG--WCENLPGSYRCTCAQGIPTRT 758  
 QY 257 -----EEDSSCY---GCTGEGGNCK--ECISGY--AREHGQCADVDECSLAEK--- 299  
 DB 759 GRLSCIDVDECEAGKYCQDQICNTPGSPQCCLSGYHLSDRSRCEDIDECDFPAACIG 818  
 QY 300 -TCVRNEN-----CYNTPGSYVVCVCPDGF 323  
 DB 819 GDCINTNGSYRCLPLGHLVGRKCKKDIDECSDPGLCLPHACENLQGSYVCVDEGF 878  
 QY 324 EETED 328  
 DB 879 TLTQD 883

RESULT 2  
 US-08-751-305-2  
 Sequence 2, Application US/08751305  
 Patent No. 5965439  
 GENERAL INFORMATION:  
 APPLICANT: Tennor et al., Andrea J.  
 TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT  
 NUMBER OF SEQUENCES: 33  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 4225 Executive Square, Suite 1400  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92037

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/751,305  
 FILING DATE: 18-NOV-1996  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Metherell, Jr., John R.  
 REGISTRATION NUMBER: 31,678  
 REFERENCE/DOCKET NUMBER: 07306/012001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619/678-5070  
 TELEFAX: 619/678-5099  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 652 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-751-305-2

Query Match 12.7%; Score 254.5; DB 2; Length 652;  
 Best Local Similarity 30.7%; Pred. No. 1.8e-12;  
 Matches 78; Conservative 27; Mismatches 88; Indels 61; Gaps 15;

QY 95 EAOEHLFAMWQLKSEYDLEFEMCVKTLKCCSPGTYGPDCLACGGSORPCS-GNGH 153  
 DB 227 EGGKDDTQSHYFLCKEKRAVDVPM-----GSSGPLCVSPKYG---CNFNNG 270  
 QY 154 CSGDGSROGDS--CRCHMGYQGPLCTDCMDGYFSSLNETHSICTACDESKTCSGLTN 211  
 DB 271 CHDDCFEGGDSGLCCRGGRFR--LDDLVY---CASRNPCCSSPPRGATCYLGHGN 325  
 QY 212 RDCGECEVGMWDEGA--CYVDCAAEPPPCSAQAQCKNANGSYTCBECSSC-VGCTG 268  
 DB 326 YTC-RCPQGYQLDSSQLDCVDVDEC--QDSPC--AQECVTPPGFRC-----CWGYEP 375  
 QY 269 EGRGNKECTSGYAREHGQCADVDECSLAEKTCVRNENCNYPGSGYVCVCPDGF----- 324  
 DB 376 GGG-----EGACQDVDECALGRSPCA--QGCTNTDGSFHSCSGEYVLAEG 420  
 QY 325 -----ETEDACVPP 333  
 DB 421 DGTQCQDVDECVCP 434

RESULT 3  
 US-08-479-722B-2  
 Sequence 2, Application US/08479722B  
 Patent No. 6074840  
 GENERAL INFORMATION:  
 APPLICANT: Bonadio, Jeffrey  
 TITLE OF INVENTION: LATENT TGF ( BINDING PROTEIN (LTBP)  
 TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Williams, Morgan & Amerson  
 STREET: 7676 Hillmont, Suite 250  
 CITY: Houston  
 STATE: Texas  
 COUNTRY: USA  
 ZIP: 77040  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:





CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patent in Release #1.0, Version  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02251  
FILING DATE: CONCURRENTLY HEREWITH  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 18-FEB-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UMIC009P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1251 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-02251-3

Query Match 12.3%; Score 246.5; DB 5; Length 1251;  
Best Local Similarity 23.5%; Pred. No. 1.5e-11;  
Matches 100; Conservative 41; Mismatches 109; Indels 175; Gaps 25;

QY 20 PAPEAAKPTPCRCRGGLVDFKNGMVDPAKKNFGGNTAMEKTLKYESSEIRLLEIL 79  
DB 517 PYBELISRPSP-----PTFRRLPLDLP-----SRSAVEIAPQTQVETDECRILNOMI 563  
QY 80 --EGLC--ESSDF--ECNOMLEAOEHLBAMWLKSEYPLDFEFVCYTKLVCCSPGY 133  
DB 564 CGHGQCVPEPSPDYSCHCNAGYRSHPOHR-----YCVDAVNE--CEAEPG 604  
QY 134 GPCLACQ--GGSQPCSGNGHCS-----GDGSRQ-----GDG----- 164  
DB 605 GPGKIGCMNTGGSY-----NCHNNGRYRLHVAGGRSCVDLNECAKPHLGDGGFCINFP 659  
QY 165 ---SCRCHNGY---QGFLCTD-----CMDGYFSSLRNETHSI--CTACDESCRTG 206  
DB 660 GHKCKCYGYRLKASRPICEDIDECRDPSTCPDG--KCEKPKGSFKCIACQPGYRSQ 716  
QY 207 SGLTRNDGCECEGVNLDGACVYDECAEAPPSCSAOFCNNANGSYTC----- 256  
DB 717 GG-----GACRDVNEC--SEGTPCSPG--WCEKLPGRSTYCTCAOQIRTRT 757  
QY 257 -----ECDSGCV---GCTGEGPGNCK--ECISGY--AREHGOCAVDDECSLAEK--- 299  
DB 758 GRLSCIDVDDCAGKVCODGICTNTPGSPFOCCLSGYHLSDRSRCEIDEDCPFAACIG 817  
QY 300 -TCVRKNEN-----GACRDVNEC--SEGTPCSPG--WCEKLPGRSTYCTCAOQIRTRT 323  
DB 818 GRLSCIDVDDCAGKVCODGICTNTPGSPFOCCLSGYHLSDRSRCEIDEDCPFAACIG 877  
QY 324 EETED 328  
DB 878 TLTGD 882

RESULT 6  
US-08-199-780-3  
Sequence 3, Application US/08199780  
Patent No. 5763416  
GENERAL INFORMATION:  
APPLICANT: Bonadio, Jeffrey  
APPLICANT: Goldstein, Steven A.  
TITLE OF INVENTION: Gene transfer into bone cells  
TITLE OF INVENTION: And tissues  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/199,780  
FILING DATE: 18-FEB-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UMIC:002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 320-7200  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1252 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-199-780-3

Query Match 12.3%; Score 246.5; DB 1; Length 1252;  
Best Local Similarity 23.5%; Pred. No. 1.5e-11;  
Matches 100; Conservative 41; Mismatches 109; Indels 175; Gaps 25;

QY 20 PAPEAAKPTPCRCRGGLVDFKNGMVDPAKKNFGGNTAMEKTLKYESSEIRLLEIL 79  
DB 518 PYBELISRPSP-----PTFRRLPLDLP-----SRSAVEIAPQTQVETDECRILNOMI 564  
QY 80 --EGLC--ESSDF--ECNOMLEAOEHLBAMWLKSEYPLDFEFVCYTKLVCCSPGY 133  
DB 518 PYBELISRPSP-----PTFRRLPLDLP-----SRSAVEIAPQTQVETDECRILNOMI 564  
QY 134 GPCLACQ--GGSQPCSGNGHCS-----GDGSRQ-----GDG----- 164  
DB 606 GHKCKCYGYRLKASRPICEDIDECRDPSTCPDG--KCEKPKGSFKCIACQPGYRSQ 660  
QY 165 ---SCRCHNGY---QGFLCTD-----CMDGYFSSLRNETHSI--CTACDESCRTG 206  
DB 661 GHKCKCYGYRLKASRPICEDIDECRDPSTCPDG--KCEKPKGSFKCIACQPGYRSQ 717  
QY 207 SGLTRNDGCECEGVNLDGACVYDECAEAPPSCSAOFCNNANGSYTC----- 256  
DB 718 GG-----GACRDVNEC--SEGTPCSPG--WCEKLPGRSTYCTCAOQIRTRT 758  
QY 257 -----ECDSGCV---GCTGEGPGNCK--ECISGY--AREHGOCAVDDECSLAEK--- 299  
DB 759 GRLSCIDVDDCAGKVCODGICTNTPGSPFOCCLSGYHLSDRSRCEIDEDCPFAACIG 818  
QY 300 -TCVRKNEN-----GACRDVNEC--SEGTPCSPG--WCEKLPGRSTYCTCAOQIRTRT 323

Db 819 GDCINTNGSYRCLCPGLHRLVGRKCKKDIDECSDPGLCLPHACENLQSGSYVCVCDGCF 878  
QY 324 EETED 328  
Db 879 TLTOD 883

## RESULT 7

US-08-316-650-3  
; Sequence 3, Application US/08316650  
; Patent No. 5942496  
; GENERAL INFORMATION:  
; APPLICANT: Bonadio, Jeffrey  
; APPLICANT: Roessler, Blake J.  
; APPLICANT: Goldstein, Steven A.  
; APPLICANT: Lin, Mushan  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS  
; TITLE OF INVENTION: FOR STIMULATING BONE CELLS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/316,650  
; FILING DATE: 30-SEP-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/199,780  
; FILING DATE: 30-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: UMIC:008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1252 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-316-650-3

Query Match 12.3% Score 246.5; DB 2; Length 1252;

Best Local Similarity 23.5% Pred. No. 1.5e-11; Mismatches 109; Indels 175; Gaps 25;

Db 20 PAPERAKPTCHRCRGVLDKFNQGMVDTAKKNFGGNTAMDEKTLKSYESSEIRLETL 79  
QY 518 PYPELISRSP-----PFHRLPLRP-----SRSAVIAITQYTEIDECLNONTI 564  
Db 80 --EGIC--ESSDF--ECNOMLEAOEHLAWMLQKSEYDLEFEMFCVYKTLKVCSPGT 133  
QY 565 CGHGQCVGPSPDYSCHCNAGYRSHQHR-----YCVAVNE--CEAEPG 605  
Db 134 GPDLACQ--GGSQRPGSGNGHCS-----GDGSRQ-----GDG----- 164  
QY 606 GPGKICIKNTGGSY-----NCHCNRGYRLHVGAGGSCVDLNECAKPHLCGGGSGCFINP 660  
QY 165 ---SCRCHMGY-----QGPLCTD-----CMDGYFSSLRNETHSI--CTACDESCCTC 206

Db 661 GHYKNCYVGYRLKASRPICEDIDECDPSTCPDG---KCENRKGSFKCIACQPGYRSQ 717  
QY 207 SGLTRNRCGCEGEVGVNIDEGACVYDDECAAEPPPCSAOFCCKNANGSTTC----- 256  
Db 718 GG-----GACRDVNEC--SEGTSPSPG--WCEKLPKSYCTCAOIGIRTRT 758  
QY 257 -----ECCDSCV---GCTGEGPCK--ECISGY--AREHGQCAVDDECSLAEK--- 299  
Db 759 GRLSCIDVDCEAGKVCODGICTNTPGSFQCCLSGHLHSDRSCEDIDECDPPACIG 818  
QY 300 -TVARKNEN-----CYNTPGSSYVCVCPDGF 323  
Db 819 GDCINTNGSYRCLCPGLHRLVGRKCKKDIDECSDPGLCLPHACENLQSGSYVCVCDGCF 878  
QY 324 EETED 328  
Db 879 TLTOD 883

## RESULT 8

5177197-30  
; Patent No. 5177197  
; APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;  
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSON-WELSH,  
; LENA; HELDIN, CARL-HENRIK  
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING  
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN  
; NUMBER OF SEQUENCES: 53  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/487,343  
; FILING DATE: 27-FEB-1990  
; SEQ ID NO:30:  
; LENGTH: 1394  
; 5177197-30

Query Match 12.0% Score 240.5; DB 6; Length 1394;

Best Local Similarity 27.1% Pred. No. 5.2e-11; Mismatches 96; Indels 111; Gaps 19;

Db 67 KYESSEIRLELEIGLCSSEDFECNOMLEAOEHLAWMLQKSEYDLEFEMFCVYKTLKV 126  
QY 655 EYCOSGYRMTQ--RGRCEDID--ECLNPSTCPDEQ-----CVN----- 688  
Db 127 CCSPEITPDDCLACQGSQRPSCGNGHC-----SGDGR--QDGSRCRCHMGY 172  
QY 689 --SPGSY--QCVPTCEGFR---GWNQCCLDVDECLEPYNVCANGDGCSNLEGSYMSCRHXG 741  
QY 173 QGPLCTDMDGYFSSLRNETHSICACDE-----SKTCSGLTRNRCGCEGEVGVN 222  
Db 742 -----TRTPDHKHCHRDIDECQOGNLCVNGCKNTEGSRFTCTGO---GYO 783  
QY 223 LD--EGACVDYDECAAEPPPCSAOFCCKNANGSYTCEGDSGCVCTEGPCKECSIG 280  
Db 784 LSAKDQCEIDIDEC--QHHLCAHQ--CRNTEGSPFC--VCDQ---CYRASGLD----- 830  
QY 281 YAREHGQCAVDDECSLAEKTCVRKNENYNTPGSYVCVCPDGF-----ETE 327  
Db 831 -----HCEIDINECLEDRKSVQR--GDCINTAGSYDCTCPDGFQDNNKTCODINECEHP 882  
QY 328 DACVPAEAETEG 341  
Db 883 GLCPGQGECLNTEG 896

RESULT 9  
US-08-282-141-4  
; Sequence 4, Application US/08282141  
; Patent No. 5538861  
; GENERAL INFORMATION:  
; APPLICANT: Schneider, Claudio  
; APPLICANT: Varum, Brian

```

: APPLICANT: Avanzi, Giancarlo
: APPLICANT: Brancolini, Claudio
: APPLICANT: Manfioletti, Guidoalberto
: TITLE OF INVENTION: Stimulating Factor for the AXL Receptor
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Amgen Inc.
: STREET: 1840 Dehaviiland Drive
: CITY: Thousand Oaks
: STATE: California
: COUNTRY: United States
: ZIP: 91320-1789
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/282,141
: FILING DATE:
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 676 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-282-141-4

```

```

Query Match          11.9%; Score 239; DB 1; Length 676;
Best Local Similarity 25.8%; Pred. No. 3.1e-11;
Matches 92; Conservative 37; Mismatches 114; Indels 114; Gaps 22;

```

```

QY 12 LPLLLLPAPPAKPPCHRCGLVDRKNOGMDTAKKNGGNTAMEEKLSTYSS 71
DB 12 LACLLLVLPVSEA-----NLLSKQASQVLYVKRR--ANSLLEETKQGNLER 56
QY 72 EIRLELEIEGCESSDFECNOMLE--AOEHLBAMWL-----OLKSEYP 114
DB 57 -----ECIEELCNKE--EAREVFENDPEPTYFPKYIYVLCNSFOTGLFTAAKQSTNAP 109
QY 115 LFEWFCVTKLKYCCSPGYGPDCLACQGSORPCSGNCHCS--GDGSRQDSCRCRMGYO 173
DB 110 LRS--CVNAIPDQCP-----LPCNEDGYMSCKDG--KASFTCTCKPQM 150
QY 174 GPLC-----TDCMDGYFSSLRNETHSICGTACDESKCTCSGLTNRDCGCEVGNVL--DEGA 227
DB 151 GERCEFDINECKD-----PSNINGGCSQICDNTPG--SYHC--SCKNGFVYMLSNKK 198
QY 228 CVDVDECAAPPPCSAOPCKNANGSYTCEDSSCVGCTGEGPNCCKECSGYA--REH 285
DB 199 CKVDDECSLKPSTICGTA--VCKNIPGDFEC-----ECPEGYRYNKK 238
QY 286 GQCADVDECSLAETKCVKRNENCYNTPGSYVCV--PDGFETED-----ACVP 332
DB 239 KSCCEDIDECSS--ENMCA---QLCVNYPGGYTCYCGKKGKFLAODQKSCCEVSVCLP 290

```

```

RESULT 10
US-08-435-434-3
: Sequence 3, Application US/08435434
: Patent No. 5714385
: GENERAL INFORMATION:
: APPLICANT: Mather, Jennie P.
: APPLICANT: Li, Ronghao
: APPLICANT: Chen, Jian
: TITLE OF INVENTION: ISOLATING AND CULTURING SCHWANN CELLS
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESS: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd

```

```

: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/435,434
: FILING DATE: 10-MAY-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Lee, Wendy M.
: REGISTRATION NUMBER: 00,000
: REFERENCE/DOCKET NUMBER: 946-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-1994
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 676 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: US-08-435-434-3

```

```

Query Match          11.9%; Score 239; DB 1; Length 676;
Best Local Similarity 25.7%; Pred. No. 3.1e-11;
Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;

```

```

QY 11 LPLLLLPAPPAKPPCHRCGLVDRKNOGMDTAKKNGGNTAMEEKLSTYSS 70
DB 11 LACLLLVLPVSEA-----NLLSKQASQVLYVKRR--ANSLLEETKQGNLER 56
QY 71 SEIRLELEIEGCESSDFECNOMLE--AOEHLBAMWL-----OLKSEYP 113
DB 57 -----ECIEELCNKE--EAREVFENDPEPTYFPKYIYVLCNSFOTGLFTAAKQSTNAP 108
QY 114 LFEWFCVTKLKYCCSPGYGPDCLACQGSORPCSGNCHCS--GDGSRQDSCRCRMGYO 172
DB 109 DLRS--CVNAIPDQCP-----LPCNEDGYMSCKDG--KASFTCTCKPQM 149
QY 173 GPLC-----TDCMDGYFSSLRNETHSICGTACDESKCTCSGLTNRDCGCEVGNVL--DEG 226
DB 150 GERCEFDINECKD-----PSNINGGCSQICDNTPG--SYHC--SCKNGFVYMLSNKK 197
QY 227 CVDVDECAAPPPCSAOPCKNANGSYTCEDSSCVGCTGEGPNCCKECSGYA--RE 284
DB 198 DCDVDECSLKPSTICGTA--VCKNIPGDFEC-----ECPEGYRYNKK 237
QY 285 HGQCADVDECSLAETKCVKRNENCYNTPGSYVCV--PDGFETED-----ACVP 332
DB 238 KSCCEDIDECSS--ENMCA---QLCVNYPGGYTCYCGKKGKFLAODQKSCCEVSVCLP 290

```

```

RESULT 11
US-08-435-436-3
: Sequence 3, Application US/08435436
: Patent No. 5721139
: GENERAL INFORMATION:
: APPLICANT: Mather, Jennie P.
: APPLICANT: Li, Ronghao
: APPLICANT: Chen, Jian
: TITLE OF INVENTION: ISOLATING AND CULTURING SCHWANN CELLS
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESS: Genentech, Inc.

```

STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: patin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/435,436  
 FILING DATE: 10-MAY-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lee, Wendy M.  
 REGISTRATION NUMBER: 00,000  
 REFERENCE/DOCKET NUMBER: 946-3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-1994  
 TELEFAX: 415/952-9881  
 TELEX: 910/371-7168  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 676 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-435-436-3

Query Match 11.9%; Score 239; DB 1; Length 676;  
 Best Local Similarity 25.7%; Pred. No. 3.1e-11;  
 Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;

QY 11 LRLPLLLPAPPAEAKKPTCHRCRGIVDFNFGMVTAKKNGGTAMKEKLSKYES 70  
 DB 11 LRLPLLLPAPPAEAKKPTCHRCRGIVDFNFGMVTAKKNGGTAMKEKLSKYES 70  
 QY 71 SEIRLEILEGLCESSDFECNQMLE--AOEHLLEAWL-----QKSEYP 113  
 DB 57 -----ECIEELCNKE--EAREVEFENDPETYKYLVLRSFOTGLFTAAQSTNAP 108  
 QY 114 DLEFMCVTKLYKCCSPGTGPDCLACQGGSQRPCSGNGHCS--GDGSRQDGSRCRMGY 172  
 DB 109 DLRS--CVNAIPDQCS-----LPCNEDGYMSCKD--KASFTCTCKPGW 149  
 QY 173 QGPLC-----TDCMDGYFSSLRNETHSICTACDESKTCSGLTNRDCECEYGVWL--DEG 226  
 DB 150 QGEKCEPDINECKD-----PSNINGGCSQICDNTPG--SYHC--SCKNGFVMSLNK 197  
 QY 227 ACVDVDECAAEPPPCSAOAFCKNANGSYTCECDSSCVGCTGEGPGNCKECISGYA--RE 284  
 DB 198 DCKVDDECSLKPSCGTA--VCKNIPDGFEC-----BCPEGYRYNLK 237  
 QY 285 HGOCADVDECSLAEKTCVRKNENYNTPGSYVCV--PDGFEETED-----ACVP 332  
 DB 238 SKSCEDIDEC--ENMCA---QLCVNYPGGTYCYCDGKKGFKLAQDOKSCEVSVCLP 290

RESULT 12  
 US-08-438-863-3  
 ; Sequence 3, Application US/08438863  
 ; Patent No. 5849585  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mather, Jennie P.  
 ; APPLICANT: Ronghao Li  
 ; APPLICANT: Chen, Jian  
 ; TITLE OF INVENTION: ISOLATING AND CULTURING SCHWANN CELLS  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: patin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/438,863  
 FILING DATE: 10-MAY-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lee, Wendy M.  
 REGISTRATION NUMBER: 00,000  
 REFERENCE/DOCKET NUMBER: 946  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-1994  
 TELEFAX: 415/952-9881  
 TELEX: 910/371-7168  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 676 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-438-863-3

Query Match 11.9%; Score 239; DB 2; Length 676;  
 Best Local Similarity 25.7%; Pred. No. 3.1e-11;  
 Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;

QY 11 LRLPLLLPAPPAEAKKPTCHRCRGIVDFNFGMVTAKKNGGTAMKEKLSKYES 70  
 DB 11 LRLPLLLPAPPAEAKKPTCHRCRGIVDFNFGMVTAKKNGGTAMKEKLSKYES 70  
 QY 71 SEIRLEILEGLCESSDFECNQMLE--AOEHLLEAWL-----QKSEYP 113  
 DB 57 -----ECIEELCNKE--EAREVEFENDPETYKYLVLRSFOTGLFTAAQSTNAP 108  
 QY 114 DLEFMCVTKLYKCCSPGTGPDCLACQGGSQRPCSGNGHCS--GDGSRQDGSRCRMGY 172  
 DB 109 DLRS--CVNAIPDQCS-----LPCNEDGYMSCKD--KASFTCTCKPGW 149  
 QY 173 QGPLC-----TDCMDGYFSSLRNETHSICTACDESKTCSGLTNRDCECEYGVWL--DEG 226  
 DB 150 QGEKCEPDINECKD-----PSNINGGCSQICDNTPG--SYHC--SCKNGFVMSLNK 197  
 QY 227 ACVDVDECAAEPPPCSAOAFCKNANGSYTCECDSSCVGCTGEGPGNCKECISGYA--RE 284  
 DB 198 DCKVDDECSLKPSCGTA--VCKNIPDGFEC-----BCPEGYRYNLK 237  
 QY 285 HGOCADVDECSLAEKTCVRKNENYNTPGSYVCV--PDGFEETED-----ACVP 332  
 DB 238 SKSCEDIDEC--ENMCA---QLCVNYPGGTYCYCDGKKGFKLAQDOKSCEVSVCLP 290

RESULT 13  
 US-08-438-864-3  
 ; Sequence 3, Application US/08438864  
 ; Patent No. 5955420  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Hammonds, R. Glenn  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Mark, Melanie R.  
 ; APPLICANT: Mather, Jennie P.

APPLICANT: Li, Ronghao  
TITLE OF INVENTION: RSE RECEPTOR ACTIVATION  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/438,864  
FILING DATE: 10-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/402253  
FILING DATE: 10-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Wendy M. Lee  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 929P1-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 676 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-438-864-3

Query Match 11.9%; Score 239; DB 2; Length 676;  
Best Local Similarity 25.7%; Pred. No. 3.1e-11;  
Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;  
QY 11 LPLLLLPAPPAKPPCHRCGLVDKFNQGVDTAKKNGGNTAMEKTLSTKYES 70  
DB 11 LACLCLLVLPVSEA-----NFLSKQASQVLVKKRR--ANSLLLETKQGNLER 56  
QY 71 SEIRLEILEIGCESSDPECNOMLE--AOEHLLEAMWL-----QLKSEYP 113  
DB 57 -----ECIEELCNKE--EAREVFENDPETDYFPRYLVCLNSFOTGLFTAROSTNATP 108  
QY 114 DLEFMCVTKLVKCSPGTYGPDCLACOGGSGRPGSGNGHCS-GDGSROGDSGRCHMGY 172  
DB 109 DLRS--CVNAIPDQSP-----LPCNEDGYMSCKRG--KASTCTCKRGW 149  
QY 173 QGPLC-----TDCMDGYFSSLRNETHSICTACDESCGTCGGLTNRDCGEGEVGWL--DEG 226  
DB 150 QGEKCEFDINECKD-----PSNINCGCSQICDNTPG--SYHC--SCKNGFVLMNKK 197  
QY 227 ACVDVDECAEPPPCSAOFCNANGSYTCEBDSSVCCTGEGGNCCKECLISGTA--RE 284  
DB 198 DCKDVDECSLKPSIGTA--VCKNIPGDFEC-----ECPEGRYNYLK 237  
QY 285 HQGCAADVDCSLAEKTCVKKNNENYNTPGSYVCV--PDGFETED-----ACVP 332  
DB 238 SKSCEDIKCS--ENMCA---OLCVNYPGTYCYCDGKKGKFLADQKSCYVSVCLP 290

RESULT 14  
US-08-438-862-3  
; Sequence 3, Application US/08438862  
; Patent No. 6033660  
; GENERAL INFORMATION:  
; APPLICANT: Mather, Jennie P.

APPLICANT: Li, Ronghao  
APPLICANT: Chen, Jian  
TITLE OF INVENTION: ISOLATING AND CULTURING SCHWANN CELLS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/438,862  
FILING DATE: 10-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 946-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 676 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-438-862-3

Query Match 11.9%; Score 239; DB 3; Length 676;  
Best Local Similarity 25.7%; Pred. No. 3.1e-11;  
Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;  
QY 11 LPLLLLPAPPAKPPCHRCGLVDKFNQGVDTAKKNGGNTAMEKTLSTKYES 70  
DB 11 LACLCLLVLPVSEA-----NFLSKQASQVLVKKRR--ANSLLLETKQGNLER 56  
QY 71 SEIRLEILEIGCESSDPECNOMLE--AOEHLLEAMWL-----QLKSEYP 113  
DB 57 -----ECIEELCNKE--EAREVFENDPETDYFPRYLVCLNSFOTGLFTAROSTNATP 108  
QY 114 DLEFMCVTKLVKCSPGTYGPDCLACOGGSGRPGSGNGHCS-GDGSROGDSGRCHMGY 172  
DB 109 DLRS--CVNAIPDQSP-----LPCNEDGYMSCKRG--KASTCTCKRGW 149  
QY 173 QGPLC-----TDCMDGYFSSLRNETHSICTACDESCGTCGGLTNRDCGEGEVGWL--DEG 226  
DB 150 QGEKCEFDINECKD-----PSNINCGCSQICDNTPG--SYHC--SCKNGFVLMNKK 197  
QY 227 ACVDVDECAEPPPCSAOFCNANGSYTCEBDSSVCCTGEGGNCCKECLISGTA--RE 284  
DB 198 DCKDVDECSLKPSIGTA--VCKNIPGDFEC-----ECPEGRYNYLK 237  
QY 285 HQGCAADVDCSLAEKTCVKKNNENYNTPGSYVCV--PDGFETED-----ACVP 332  
DB 238 SKSCEDIKCS--ENMCA---OLCVNYPGTYCYCDGKKGKFLADQKSCYVSVCLP 290

RESULT 15  
US-08-628-747-3  
; Sequence 3, Application US/08628747  
; Patent No. 6169070  
; GENERAL INFORMATION:  
; APPLICANT: Mather, Jennie P.

APPLICANT: Chen, Jian  
APPLICANT: Godowski, Paul J.  
APPLICANT: Hammonds, R. Glenn  
APPLICANT: Mark, Melanie  
APPLICANT: Mather, Jennie P.  
APPLICANT: Li, Ronghao  
TITLE OF INVENTION: RECEPTOR ACTIVATION BY GAS6  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: One DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/628,747  
FILING DATE: 17-APR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/402,253  
FILING DATE: 10-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/438,861  
FILING DATE: 10-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: P929P2PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-225-1994  
TELEFAX: 650-952-9881  
TELEX: 910-371-7168  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 676 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-628-747-3

Query Match 11.9%; Score 239; DB 4; Length 676;  
Best Local Similarity 25.7%; Pred. No. 3,1e-11;  
Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;  
QY 11 LPLLLLPPEAAKKTPTCHRCGLVDKFNQGNVDTAKKFGGNTAMEEKLTKSYES 70  
DB 11 LPLLLLPPEAAKKTPTCHRCGLVDKFNQGNVDTAKKFGGNTAMEEKLTKSYES 70  
QY 71 SEIRLEITLEGICSSDFECNOMLE-AQEHLEAMWL-----QKSEYP 113  
DB 57 -----ECIEELCKNE--EAREVFENDEPDTDFYFKYLVLCLRSFOTGLFTARQSTNAYP 108  
QY 114 DLFEWFCYKTLKVCSPPTGYPDCLACGGSGRPSGNGHCS-GDGSROGDSCCHMGY 172  
DB 109 DLRS--CVNALPDOCSP-----LPCNEDGYMSCKDG--KASTCTCKPKGW 149  
QY 173 QGPLC-----TTCMDGYFSLRNETHSICTACDESKCTCSGLTNRDCCGEVGVWL--DEG 226  
DB 150 QGEKEFPDINECKD-----PSNNGGCSQICDNTPG--SYHC-SCKNGFVWLSMK 197  
QY 227 ACVDVDECAEPPPCSAOFCNANGSYTCCEBDCSSCVGCTGEGPCKGCKECSGYA--RE 284  
DB 198 DCKDVDECSLKRPSICGTA--VCKNIPGDPEC-----ECPEGYRYNLK 237  
QY 285 HGOCAADVDECSLAERTCVARKNENCYNTPGSYVCVC--PDGFEETED-----ACVP 332

DB 238 SKSCEDIDCS--ENMCA---QLCVNYPGGYTCYCDGKKGFKRLAQDQKSCFEVSVCLP 290  
RESULT 16  
US-08-402-253-3  
Sequence 3, Application US/08402253  
Patent No. 6211142  
GENERAL INFORMATION:  
APPLICANT: Chen, Jian  
APPLICANT: Hammonds, R. Glenn  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
TITLE OF INVENTION: RSE RECEPTOR ACTIVATION  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 KB floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/402,253  
FILING DATE: 10-MAR-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wendy M. Lee  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 929  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 676 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-402-253-3

Query Match 11.9%; Score 239; DB 4; Length 676;  
Best Local Similarity 25.7%; Pred. No. 3,1e-11;  
Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;  
QY 11 LPLLLLPPEAAKKTPTCHRCGLVDKFNQGNVDTAKKFGGNTAMEEKLTKSYES 70  
DB 11 LPLLLLPPEAAKKTPTCHRCGLVDKFNQGNVDTAKKFGGNTAMEEKLTKSYES 70  
QY 71 SEIRLEITLEGICSSDFECNOMLE-AQEHLEAMWL-----QKSEYP 113  
DB 57 -----ECIEELCKNE--EAREVFENDEPDTDFYFKYLVLCLRSFOTGLFTARQSTNAYP 108  
QY 114 DLFEWFCYKTLKVCSPPTGYPDCLACGGSGRPSGNGHCS-GDGSROGDSCCHMGY 172  
DB 109 DLRS--CVNALPDOCSP-----LPCNEDGYMSCKDG--KASTCTCKPKGW 149  
QY 173 QGPLC-----TTCMDGYFSLRNETHSICTACDESKCTCSGLTNRDCCGEVGVWL--DEG 226  
DB 150 QGEKEFPDINECKD-----PSNNGGCSQICDNTPG--SYHC-SCKNGFVWLSMK 197  
QY 227 ACVDVDECAEPPPCSAOFCNANGSYTCCEBDCSSCVGCTGEGPCKGCKECSGYA--RE 284  
DB 198 DCKDVDECSLKRPSICGTA--VCKNIPGDPEC-----ECPEGYRYNLK 237





Db 651 SRCVSCEDGRYFNGOCOPCHRCATCAGAGADGCTCTEGYFMEDGRVQ----- 702  
OY 239 PRCSAOFCNA--NGSYTCECDSSCVGCTGEGPNCKECTISYAREHQ-----CAD 290  
Db 703 -SCSISTYFHDSSNGKSKCKDICTCNGPFPKNTSCPSGLDLDLGMCGAICKD 761  
OY 291 VDECSLAEK--TCVRKNENCYNTPGSYVC 317  
Db 762 ATEESMAEGGFCMLVKNNLQORRVLOQLC 791

RESULT 19  
US-08-976-838-23  
Sequence 23, Application US/08976838  
Patent No. 5981259  
GENERAL INFORMATION:  
APPLICANT: Franzusoff, Alex  
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/976,838  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2848-11-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 799 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-976-838-23

Query Match 11.8%; Score 236; DB 2; Length 799;  
Best Local Similarity 25.8%; Pred. No. 6.3e-11;  
Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;  
OY 38 VDKRNGMVTAKNFGGNTAMEKTLSESEIRLETLBLEGCSSEDFECNMLEAQ 97  
Db 498 VERFRYSRVEDPTDYGTEYAGP---CDEPCESEV-----CCDGPDPCHNDCLH-- 544  
OY 98 EHLHAWMLQKSEYPLDFEMFCVKT-----KVC--CSP-----GTGPPDLAC 140  
Db 545 -----YYKLKNN-----TRICVSCPPGHYHADKRCRCACANCSGFSHGDQCSG 593  
OY 141 QGG--SQRPCSGNGHSGDSSROG--DGSCR-----CHMGYQPLCTDCMDGYFSSLRN 190  
Db 594 KYGYFLNETNSCVTHCP--DSYQDTKKNLCRKCSKNCTCTERHNCTECRDGL--SLQG 650  
OY 191 ETHSI-----CTACDESKTCSGLTNDGCEGEVGVWLDGACVYVDECAAP 238  
Db 651 SRCVSCEDGRYFNGOCOPCHRCATCAGAGADGCTCTEGYFMEDGRVQ----- 702

OY 239 PRCSAOFCNA--NGSYTCECDSSCVGCTGEGPNCKECTISYAREHQ-----CAD 290  
Db 703 -SCSISTYFHDSSNGKSKCKDICTCNGPFPKNTSCPSGLDLDLGMCGAICKD 761  
OY 291 VDECSLAEK--TCVRKNENCYNTPGSYVC 317  
Db 762 ATEESMAEGGFCMLVKNNLQORRVLOQLC 791

RESULT 20  
US-08-525-940-21  
Sequence 21, Application US/08525940  
Patent No. 5866351  
GENERAL INFORMATION:  
APPLICANT: Franzusoff, Alex  
APPLICANT: Miranda, Luis R.  
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES  
TITLE OF INVENTION: ENCODING SAID PROTEASES  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/525,940  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/368,852  
FILING DATE: 01-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/088,322  
FILING DATE: 07-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2848-11-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 881 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-525-940-21

Query Match 11.8%; Score 236; DB 2; Length 881;  
Best Local Similarity 25.8%; Pred. No. 7.1e-11;  
Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;  
OY 38 VDKRNGMVTAKNFGGNTAMEKTLSESEIRLETLBLEGCSSEDFECNMLEAQ 97  
Db 580 VERFRYSRVEDPTDYGTEYAGP---CDEPCESEV-----CCDGPDPCHNDCLH-- 626  
OY 98 EHLHAWMLQKSEYPLDFEMFCVKT-----KVC--CSP-----GTGPPDLAC 140  
Db 627 -----YYKLKNN-----TRICVSCPPGHYHADKRCRCACANCSGFSHGDQCSG 675  
OY 141 QGG--SQRPCSGNGHSGDSSROG--DGSCR-----CHMGYQPLCTDCMDGYFSSLRN 190  
Db 676 KYGYFLNETNSCVTHCP--DSYQDTKKNLCRKCSKNCTCTERHNCTECRDGL--SLQG 732



Db 710 KYGFLENETNSCVTHCP-DGSYODTKKNNLCRKCSNCKTCTEFHNCYECRDL--SLQG 766  
 QY 191 ETHSI-----CTADESCKTCGSLTNRDCCGECEVGMVDEGACVDECAEP 238  
 Db 767 SRCVSCEDEGRYFNQDQPCPHRCATCAGAGADGCINCTEGYFMEDERCVQ----- 818  
 QY 239 PRCSAOFCCKNA--NGSYTCECDSSCVGCTGEGPNCCKECISGYAREHGO-----CAD 290  
 Db 819 -SCSISYFHDHSSSENGYKSKCKDCISLTCNGPGRKNTSCPSGLDLGMCQMGATCKD 877  
 QY 291 VDECSLAER---TCVRKNENCTNTPGSYVC 317  
 Db 878 ATEESMAEGGFCMLVKNNLCKRKVLDQLC 907

## RESULT 23

US-08-976-838-18  
 ; Sequence 18, Application US/08976838  
 ; Patent No. 5981259  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Franzusoff, Alex  
 ; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID  
 ; TITLE OF INVENTION: MOLECULES  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sheridan Ross P.C.  
 ; STREET: 1700 Lincoln St., Suite 3500  
 ; CITY: Denver  
 ; STATE: Colorado  
 ; COUNTRY: U.S.A.  
 ; ZIP: 80203  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/976,838  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Connell, Gary J.  
 ; REGISTRATION NUMBER: 32,020  
 ; REFERENCE/DOCKET NUMBER: 2848-11-C2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (303) 863-9700  
 ; TELEFAX: (303) 863-0223  
 ; INFORMATION FOR SEQ ID NO: 18:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 915 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-976-838-18

Query Match 11.8%; Score 236; DB 2; Length 915;  
 Best Local Similarity 25.8%; Pred. No. 7.4e-11;  
 Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;

QY 38 VDKFNQGVADPAKKNFGGGNTAMEEKTLSKYESSIRLEILEGSCSSDFECNOMLEAQ 97  
 Db 614 VERRRYSEVDEPTDDYDGEDYAGP-----CDPECSFV-----GDDGFGPDHCNDCLH-- 660  
 QY 98 EHLLEAMVQLAKSEYPDLEFMEVCVTKL-----KYC--CSP-----GTGPDCLAC 140  
 Db 661 -----YYKKLKN-----TRICVSSCPGHYHADKRCRCAPNCESCFCGSHGQCMSC 709  
 QY 141 OGC---SQRPSGNGCHGSGDGRG--DGSCR-----CHMGYQGLCTDCMDGYFSSLRN 190  
 Db 710 KYGFLENETNSCVTHCP-DGSYODTKKNNLCRKCSNCKTCTEFHNCYECRDL--SLQG 766

QY 191 ETHSI-----CTADESCKTCGSLTNRDCCGECEVGMVDEGACVDECAEP 238  
 Db 767 SRCVSCEDEGRYFNQDQPCPHRCATCAGAGADGCINCTEGYFMEDERCVQ----- 818  
 QY 239 PRCSAOFCCKNA--NGSYTCECDSSCVGCTGEGPNCCKECISGYAREHGO-----CAD 290  
 Db 819 -SCSISYFHDHSSSENGYKSKCKDCISLTCNGPGRKNTSCPSGLDLGMCQMGATCKD 877  
 QY 291 VDECSLAER---TCVRKNENCTNTPGSYVC 317  
 Db 878 ATEESMAEGGFCMLVKNNLCKRKVLDQLC 907

## RESULT 24

US-08-525-940-15  
 ; Sequence 15, Application US/08525940  
 ; Patent No. 5866351  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Franzusoff, Alex  
 ; APPLICANT: Miranda, Luis R.  
 ; APPLICANT: Wolf, Joseph R.  
 ; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES  
 ; TITLE OF INVENTION: ENCODING SAID PROTEASES  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sheridan Ross & McIntosh  
 ; STREET: 1700 Lincoln Street, Suite 3500  
 ; CITY: Denver  
 ; STATE: Colorado  
 ; COUNTRY: U.S.A.  
 ; ZIP: 80203  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/525,940  
 ; FILING DATE:  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/368,852  
 ; FILING DATE: 01-JAN-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/088,322  
 ; FILING DATE: 07-JUL-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Connell, Gary J.  
 ; REGISTRATION NUMBER: 32,020  
 ; REFERENCE/DOCKET NUMBER: 2848-11-C1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (303) 863-9700  
 ; TELEFAX: (303) 863-0223  
 ; INFORMATION FOR SEQ ID NO: 15:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 288 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-525-940-15

Query Match 11.5%; Score 230.5; DB 2; Length 288;  
 Best Local Similarity 28.8%; Pred. No. 5.6e-11;  
 Matches 66; Conservative 31; Mismatches 81; Indels 51; Gaps 12;

QY 128 CSP-----GTGPDCLACOG---SORPSGNGCHGSGDGRG--DGSCR-CHMGY--- 172  
 Db 64 CAPNCESGFGSHGDDCMGCKRIGYFLNETNSCVTHCP-DGSYODTKKNNLCRKCSNFKTC 122  
 QY 173 -QGPLCTDCMDGYFSSLRNETHSI-----CTADESCKTCGSLTNRDCCGECEV 219  
 Db 123 TEFHICTECRDL--SLQSGRCSVCEDEGRYFIGDQPCPHRFATCAGAGADGCINCTE 180

```

OY 220 GWLDEGCACVYDDECAAEPPSCSAOACCKNA--NGSYTCEQDSYCGYCGEEGNGKEC 277
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 181 GFYFEDGRCVGI-----CSTSYTFDHSSENGYKSCCKDCDLSCLTLCNGPGRKNTSC 231

OY 278 ISGTAREHGO-----CADVDECSLAEK--TCYRKENCYNTPGSYVC 317
    |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 232 PSGILLIDGKCGMCAICADATETESWARCGFCMLVKKNNLCCORRYLQDQ 280

```

## RESULT 25

```

US-08-976-838-15
: Sequence 15, Application us/08976838
: Patent No. 5981259
:
: GENERAL INFORMATION:
: APPLICANT: Franzusoff, Alex
: TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
: TITLE OF INVENTION: MOLECULES
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Ross P.C.
: STREET: 1700 Lincoln St., Suite 3500
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.A.
: ZIP: 80203
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/976,838
: FILING DATE:
:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2846-11-C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 288 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-976-838-15

```

Query Match	11.5%	Score 230.5	DB 2	Length 288
Best Local Similarity	28.8%	Pred. No. 5.6e-11		
Matches 66	Conservative 31	Mismatches 81	Indels 51	Gaps 12

QY	128	CSP-----GTYGPDCLAACOG---SQRPSGNGHGSGDSRQ--DGSQR-CHMGV---	172
	1		
	2		
	3		
	4		
	5		
	6		
	7		
	8		
	9		
	10		
	11		
	12		
	13		
	14		
	15		
	16		
	17		
	18		
	19		
	20		
	21		
	22		
	23		
	24		
	25		
	26		
	27		
	28		
	29		
	30		
	31		
	32		
	33		
	34		
	35		
	36		
	37		
	38		
	39		
	40		
	41		
	42		
	43		
	44		
	45		
	46		
	47		
	48		
	49		
	50		
	51		
	52		
	53		
	54		
	55		
	56		
	57		
	58		
	59		
	60		
	61		
	62		
	63		
	64		
	65		
	66		
	67		
	68		
	69		
	70		
	71		
	72		
	73		
	74		
	75		
	76		
	77		
	78		
	79		
	80		
	81		
	82		
	83		
	84		
	85		
	86		
	87		
	88		
	89		
	90		
	91		
	92		
	93		
	94		
	95		
	96		
	97		
	98		
	99		
	100		
	101		
	102		
	103		
	104		
	105		
	106		
	107		
	108		
	109		
	110		
	111		
	112		
	113		
	114		
	115		
	116		
	117		
	118		
	119		
	120		
	121		
	122		
	123		
	124		
	125		
	126		
	127		
	128		
	129		

## RESULT 26

US-07-907-190-1

Sequence 1 Application US/07907190  
Patent No. 5321123  
GENERAL INFORMATION:  
APPLICANT: Griffin, John H.  
APPLICANT: Fernandez, Jose A.  
TITLE OF INVENTION: PROTEIN S POLYPEPTIDES AND ANTI-PEPTIDE  
TITLE OF INVENTION: ANTIBODIES THAT INHIBIT PROTEIN S BINDING TO C4b BINDING  
TITLE OF INVENTION: PROTEIN, DIAGNOSTIC SYSTEMS AND THERAPEUTIC METHODS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: The Scripps Research Institute, Office of  
ADDRESSSEE: Patent Counsel  
STREET: 10666 No. 5321123th Torrey Pines Road, Mail Drop TPC8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/907,190  
FILING DATE: 19920701  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/724,746  
FILING DATE: 02-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: SCRI119P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-554-2937  
TELEFAX: 619-554-6312  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 635 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-07-907-190-1

Query Match	11.3%	Score 227.5;	DB 1;	Length 635;
Best Local Similarity	28.6%	Pred. NO. 2.3e-10;		
Matches 70;	Conservative 26;	Mismatches 74;	Indels 75;	Gaps 16;

Qy	107	QKSKSEYPLDFEMFCVKTLLKCCSGPYGPDLCQGGSSORPCSGNHCS--GDSRQGDG	155
Db	61	OSTNATYPLPLRS--CYNAIPLDCCS-----LPCNIDGYSMCKDG--KASFT	101
Qy	166	CRCHMGVYOGPLC---TDKMDGYFSSLNETHSICTACDESKCTSGLTINRDGCECEVGM	221
Db	102	CTCKPGMGEGCEPDIINCKD-----PSNINGGCSOICDITNPG--SYHC-SCKNGF	149
Qy	222	VL--DEGACVYVDDCAAEPPPCSAAGPCFKNANGSYTDEEDSSCVGCTGGBGPNCKRCIS	279
Db	150	VMLSNKKCKCKVDECSLKPSICGTA-VCKNIPLGDFEC-----ECPE	189
Qy	280	GYA--REHGQCADVDECSLAERTCVRRNENNCYNPGSYVCV--PDGFEEED-----	338
Db	190	GYYRLTKSKSCEDIDECSS--ENMCA---QLCVNYPGGYCTCDGKKGFKLAQDQKSCSEV	244
Qy	329	-ACVP	332
Db	245	SVCLP	249

## RESULT 27

US-07-985-691-2

Sequence 2, Application US/07985691  
Patent No. 5405946  
GENERAL INFORMATION:  
APPLICANT: Griffith, John H  
APPLICANT: Bouma, Bonno N  
APPLICANT: Bertina, Rogier  
TITLE OF INVENTION: RECOMBINANT PROTEIN S VARIANTS DEFICIENT  
IN CABP BINDING ACTIVITY, COMPOSITIONS AND THERAPEUTIC  
METHODS  
TITLE OF INVENTION: IN CABP BINDING ACTIVITY, COMPOSITIONS AND THERAPEUTIC  
METHODS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10666 No. 5405946th Torrey Pines Road, TPC 8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/985,691  
FILING DATE: 19921202  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSP0042P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-554-2937  
TELEFAX: 619-554-6312  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 635 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-985-691-2

Query Match 11.3%; Score 227.5; DB 1; Length 635;  
Best Local Similarity 28.6%; Pred. No. 2.3e-10;  
Matches 70; Conservative 26; Mismatches 74; Indels 75; Gaps 16;  
QY 107 QLKSEYDPLFEMFCVKTLYKCCSPGTGPPDCLACGGSGRPGSGNGHCS-GDGSROGDGS 165  
Db 61 QSTNAYPDLRS--CVNAIPDQCS-----LPCNEDGYMSCKDG--KASFT 101  
QY 166 CRCHMGYOGPLC---TDCMGYFSSLRNETHSICTACDESCRTCSGLTNRDCGCEYGVW 221  
Db 102 CTCKRGMGEGCEPDIINCKD-----PSNINGGGSQICDNTPG--SYHC-SCKNGF 149  
QY 222 VL--DEGACVDVDECAAEPPCSAQAQFCKNANGSYTCECCSSCVGCTGEBGPNCKECIS 279  
Db 150 VMLNKKDKCKYDDECSLKPSTICGTA-YCKNIPGDEFC-----ECPE 189  
QY 280 GYA--REHGOCADVDECSLAERTCVRNKNENYTPGYSVVC--PDGFEETED----- 328  
Db 190 GYRNLKSKSCEDIDEC--ENMCA---QLCVNYPGTYCYCDGKKGFKLADQKSCREV 244  
QY 329 -ACVP 332  
Db 245 SVCLP 249

RESULT 28  
US-08-436-804-2  
Sequence 2, Application US/08436804  
Patent No. 5656484  
GENERAL INFORMATION:

APPLICANT: BOUMA, Bonno N.  
APPLICANT: BERTINA, Rogier M.  
TITLE OF INVENTION: PROTEIN S DELETION VARIANTS DEFICIENT IN  
CABP BINDING ACTIVITY, BUT HAVING APC COFACTOR ACTIVITY,  
COMPOSITIONS AND THERAPEUTIC METHODS  
TITLE OF INVENTION: CABP BINDING ACTIVITY, BUT HAVING APC COFACTOR ACTIVITY,  
COMPOSITIONS AND THERAPEUTIC METHODS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Young & Thompson  
STREET: Second floor, 745 South 23rd Street  
CITY: Arlington  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,804  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/267,387  
FILING DATE: 29-JUN-1994  
APPLICATION NUMBER: EP 93 201 906.0  
FILING DATE: 30-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: B038633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 635 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-436-804-2

Query Match 11.3%; Score 227.5; DB 1; Length 635;  
Best Local Similarity 28.6%; Pred. No. 2.3e-10;  
Matches 70; Conservative 26; Mismatches 74; Indels 75; Gaps 16;  
QY 107 QLKSEYDPLFEMFCVKTLYKCCSPGTGPPDCLACGGSGRPGSGNGHCS-GDGSROGDGS 165  
Db 61 QSTNAYPDLRS--CVNAIPDQCS-----LPCNEDGYMSCKDG--KASFT 101  
QY 166 CRCHMGYOGPLC---TDCMGYFSSLRNETHSICTACDESCRTCSGLTNRDCGCEYGVW 221  
Db 102 CTCKRGMGEGCEPDIINCKD-----PSNINGGGSQICDNTPG--SYHC-SCKNGF 149  
QY 222 VL--DEGACVDVDECAAEPPCSAQAQFCKNANGSYTCECCSSCVGCTGEBGPNCKECIS 279  
Db 150 VMLNKKDKCKYDDECSLKPSTICGTA-YCKNIPGDEFC-----ECPE 189  
QY 280 GYA--REHGOCADVDECSLAERTCVRNKNENYTPGYSVVC--PDGFEETED----- 328  
Db 190 GYRNLKSKSCEDIDEC--ENMCA---QLCVNYPGTYCYCDGKKGFKLADQKSCREV 244  
QY 329 -ACVP 332  
Db 245 SVCLP 249

RESULT 29  
US-08-267-387-2  
Sequence 2, Application US/08267387

```

; Patent No. 5663142
; GENERAL INFORMATION:
; APPLICANT: BOUMA, Bonno N.
; APPLICANT: BERTINA, Rogier M.
; TITLE OF INVENTION: PROTEIN S DELETION VARIANTS DEFICIENT IN
; TITLE OF INVENTION: C4BP BINDING ACTIVITY, BUT HAVING APC COFACTOR ACTIVITY,
; TITLE OF INVENTION: COMPOSITIONS AND THERAPEUTIC METHODS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Young & Thompson
; STREET: Second Floor, 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,387
; FILING DATE: 29-JUN-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: EP 93 201 906.0
; FILING DATE: 30-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: B038633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-267-387-2

```

```

Query Match 11.3%; Score 227.5; DB 1; Length 635;
Best Local Similarity 28.6%; Pred. No. 2.3e-10;
Matches 70; Conservative 26; Mismatches 74; Indels 75; Gaps 16;

QY 107 QLKSEYPDLEFMEFCVTKLKVCCSPGTGPDCLACGGGSGRQPCSGNCHCS-GDSSRQGDG 165
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 QSTNAVDPDLS-CVNAPIDQCS-----LPCNEDGYMSCKDQ--KASFT 101

QY 166 CRCHMYOGPLC-----TDCMDGYFSSLRNETHSICTACDESCCTGCGLNRRDCEVEVGW 221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 102 CCKCKPMDKEKEFDINEKD-----PSNNGGCSQICDNTFG--SYHC-SCKNPF 149

QY 222 VL--DEGACVDVDECAEPPPCSAOFCNNANGSYTCEEDSSCVCTGTGPGNCKECS 279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 VMLSNKKDKDDECLSKSIGTA-VCKNIPDFEC-----ECPE 189

QY 280 GYA--REHGOCADVDECSLAETKVRKNENYNTPGSYVCV--PDGFETED----- 328
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 190 GYRYNLSKSKCEDIDCS--ENMCA---QLCVNYPGTYCYCDGKGFKLADQKSCYEV 244

QY 329 -ACVP 332
      |||
DB 245 SVCLP 249

```

RESULT 30  
5258288-4  
; Patent No. 5258288

```

; APPLICANT: Wydro, Robert; Cohen, Edward; Dackowski, William
; Stenflo, John; Lundwall, Åke; Dahlback, Björn
; TITLE OF INVENTION: VECTOR CONTAINING DNA ENCODING MATURE
; HUMAN PROTEIN S
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/890,401
; FILING DATE: 25-JUL-1986
; SEQ ID NO: 4:
; LENGTH: 652
; US-08-267-387-2

```

```

Query Match 11.2%; Score 224; DB 6; Length 652;
Best Local Similarity 25.9%; Pred. No. 4.5e-10;
Matches 83; Conservative 39; Mismatches 90; Indels 108; Gaps 20;

QY 65 LSKYESSEI-----RLLETLGLCESS-DPECNOMLEAOEHLKEM-----WL 106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 LSKQASQVLYVRKRRADSLSLETKGNTLRECEIEELCNKEAREVEFENDPETYPKYL 60

QY 107 -----OLKSEYPDLEFMEFCVTKLKVCCSPGTGPDCLACGGGSGRQPCSGN 151
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 VCLRSFOTGLFTAROSTNAVDPDLS-CVNAPIDQCS-----LPCNED 103

QY 152 GHCS-GDSSRQGDGSCRCMHGYOGPLC-----TDCMDGYFSSLRNETHSICTACDESCCTK 206
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 104 GYMSCKDQ--KASFTCTCRPMNGEKECEPDINECKD-----PSNNGGCSQICDNT 152

QY 207 SGLTNRDCEGEYGVYL--DEGACVDVDECAEPPPCSAOFCNNANGSYTCEEDSSCV 264
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 153 PG--SYHC-SCKNGFVMLSNKKDKDDECLSKSIGTA-VCKNIPDFEC----- 200

QY 265 GCTGEGPKNCKECSGYA--REHGOCADVDECSLAETKVRKNENYNTPGSYVCV--P 320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 201 -----ECPEGYRNLSKSKCEDIDCS--ENMCA---QLCVNYPGTYCYCDGK 244

QY 321 DGFETED-----ACVP 332
      ||| |||
DB 245 KGFKLADQKSCYEVSVCLP 264

```

```

RESULT 31
US-08-882-046-5
; Sequence 5, Application US/08882046
; Patent No. 6136952
; GENERAL INFORMATION:
; APPLICANT: Li, Linheng
; APPLICANT: Hood, Leroy
; APPLICANT: Krantz, Ian D.
; APPLICANT: Spinner, Nancy B.
; TITLE OF INVENTION: Human Jagged Polypeptide, Encoding
; TITLE OF INVENTION: Nucleic Acids and Methods of Use
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,046
; FILING DATE: 25-JUN-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815

```

```

: REFERENCE/DOCKET NUMBER: P-UW 2637
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (619) 535-9001
:
: TELEFAX: (619) 535-8949
:
: INFORMATION FOR SEQ ID NO: 5:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 1219 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08-882-046-5

```

Query Match	11.28;	Score 224;	DB 4;	Length 1219;
Best Local Similarity	22.88;	Pred. No. 8.9e-10;		
Matches 97; Conservative	42;	Mismatches 101;	Indels 186;	Gaps 28;

Qy	42	NOGWVD-----TAKKNFGGANTAMEEKTLSTKYSSESLRLEILEGCESS--DPECNOM	93
Db	157	HSGMINPSRQWQTLKQNTG-----IAHFE-YQIRV-----YCDHHYGFQGNKF	199
Qy	94	LEAOEHLAAMWLQKSEYDPLF-EMFCVYTKLVKCCSPGYTGPDG--LACGGSGQRPCSG	156
Db	200	CRPRD-----DFFGHYACDQNGNKTCKEKGWMPGPCNKAICRQG----CSP	240
Qy	151	NGHSGDSSRGDSCRCRCHMGYGRPLCTDGN--DGYTSSLRNE-----	191
Db	241	K-----HESCKLPDCKRQYGMGLYCDKCTLRPHGVYHGTCNEPMQCLCTETNMGGQLCDK	295
Qy	192	-----THSIC-----TACDE-----SCK-TC	206
Db	296	DLYNGTHQPLNLNGTCSNNGPDKYQCSGCPBGYSGPNCETAEAHACLSDPCHNRSCRETS	355
Qy	207	SGLTNRDGCCEVEGVN-----LDE-----GAC-----V	229
Db	356	SGF---EC-ECSPTWGTPTCSNTINDDCSPNNCSHGHTCDPLVNGFKVCVCPQWGTQTCL	411
Qy	230	DVDECAAPRPSAASFCKNANGSYTC-----EBCDSSCVCTG--EGPGNCKE--	276
Db	412	DANECEAK--PCVYARSKNLIJASTYCDCLPBMGMGNCNDININICLQCDQNDASCRDLVN	465
Qy	277	-----CISGVAREHGQCADVDGCSLAETKYCRKNENCTNTPGSYVVCYCPDGF-----EET	326
Db	470	GYRCICPGPYAGDHCE-RDIDEC--ASNPC-L-NGEHCQNEINRQCICPTGFSNLCOLD	525
Qy	327	EDACVP	332
Db	526	IDYCEP	531

RESULT 32  
 US-08-368-852-15  
 Sequence 15, Application US/08368852  
 Patent No. 5691183  
 GENERAL INFORMATION:  
 APPLICANT: Franzusoff, Alex  
 APPLICANT: Miranda, Luis R.  
 TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES  
 TITLE OF INVENTION: ENCODING SAID PROTEASES  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Sheridan Ross & McIntosh  
 STREET: 1700 Lincoln Street, Suite 3500  
 CITY: Denver  
 STATE: CO  
 COUNTRY: U.S.A.  
 ZIP: 80203  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

```

1      APPLICATION NUMBER:  US/08/368,852
2      FILING DATE:  05-JAN-1995
3      CLASSIFICATION:  435
4      ATTORNEY/AGENT INFORMATION:
5      NAME:  Verser, Carol Talkington
6      REGISTRATION NUMBER:  37,459
7      REFERENCE/DOCKET NUMBER:  2848-11
8      TELECOMMUNICATION INFORMATION:
9      TELEPHONE:  303/863-9700
10     TELEFAX:  303/863-0223
11     INFORMATION FOR SEQ ID NO:  15:
12     SEQUENCE CHARACTERISTICS:
13     LENGTH:  288 amino acids
14     TYPE:  amino acid
15     TOPOLOGY:  linear
16     MOLECULE TYPE:  protein
17
18     US-08-368-852-15

```

Query Match	11.1%	Score 223.5	DB 1	Length 288
Best Local Similarity	27.9%	Pred. NO. 2e-10		
Matches 64, Conservative	28	Mismatches 86	Indels 51	Gaps 10

```

OY 128 CSP-----GTYGPPCLACOG---SQAPCGNGHCSDSRGQDGSRCMHGTGQ---174
Db 64 CAPNESCFSHSGDDCMCKKGYFLNETNCSYVTHCP--DG5YDPTKKMLKCKSENKTC 122
OY 175 ---PLCTCDMGDYFSSLNETHSI-----CTACDSCCKTCSGLNRDRCGEV 219
Db 123 TEFHACTECRBL--SLQSCRCYSCEBGRFFXGDDCPCHRXATCAGAGADGCTCTE 180
OY 220 GWNLDGACVYDDECAEPSPCSAAOFCKNA--NGSYTCEBDSSCVGCTGEGGNCKEC 277
Db 181 GFYMGDGRGVQX-----CSISYTYDHSSENGYSCKKCDISCLTCNPGFKNCTSC 231
OY 278 ISGYAREHQ-----CADVDECSLAEK---TCVRKNENCNFTGSGSYVC 317
Db 232 PGGYLLDGMCGMGAICTKDATTEESAAEGFCMLVKKNNLCKRKVYLQIC 280

```

RESULT 33  
 US-08-185-432-18  
 ; Sequence 18 Application US/08185432  
 ; Patent No. 5750552  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Artavanis-Tsakonas, Spyridon  
 ; APPLICANT: Bussseau, Isabelle  
 ; APPLICANT: Diederich, Robert J.  
 ; APPLICANT: Xu, Tian  
 ; APPLICANT: Matsuno, Kenji  
 ; TITLE OF INVENTION: DELETED PROTEINS, NUCLEIC ACIDS, AND  
 ; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
 ; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentln Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/185,432  
 ; FILING DATE: 21-JAN-1994  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mistrock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 7326-006

```

? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 790-9090
?
? TELEFAX: (212) 869-8864/97411
?
? TELE: 66141 PENNIE
?
? INFORMATION FOR SEQ ID NO: 18:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 2523 amino acids
?
? type: amino acid
?
? TOPOLOGY: unknown
?
? MOLECULE TYPE: protein
?
US-08-185-432-18

```

Query Match	11.1%	Score 223	DB 1	Length 2523
Best Local Similarity	23.58%	Pred. No. 2.4e-09		
Matches 68	Conservative 20	Mismatches 101	Indels 100	Gaps 14

OY	126	VCSPGTYGPDL-----ACGGSGRRCSCGSHCSGSGSRQGSQCSCHMGYGPLCTD-	179
Db	164	ICKCPFGHGATCKODLINEC----SQNPKNGGCCLINE---FGSYRCTQNRFGRNCDEP	217
OY	180	-----CMDGYSSLRNETHSICTA-----CDSEKTC-----SGLTN	211
Db	218	YVPCNPSCPILNGGTCRQTDIDTTSYDCTCLPGFSGONCEENIDDCPSNNCRNGGFCVDGVNT	277
OY	212	RDCGECEVGWLDSGACVYDVECAAEPPPPCSAOFPCANANGSTYC-----EECDSS-	262
Db	278	YN-CQCPDMT--GGYTCEDEDVCOLMFPACONGSTCENTYGGIVCYCVNGMTGEDCSENT	335
OY	263	-----CWGCTGEGBNKE-----CISGARERHQC-----	288
Db	336	DDCANACHSGATCHDRVASFYECBPHRGRTGLCLLDNACISNPCNCGSNCDITNPVNGKA	395
OY	289	-----ADVDEGLAKTKTVRRKNENCYNTPGSIYVCYCPGPF	323
Db	396	ICTGCPGTYGPACNNNDVECSLGNAPC-EHGGRTNLTLGSCQCCPQGY	443

```

RESULT 34
US-09-110-116-3
Sequence 3, Application US/09110116
Patent No. 6013479
GENERAL INFORMATION:
APPLICANT: Xu, Hong
APPLICANT: Cohen, Victoria L.
APPLICANT: Stuart, Susan G.
TITLE OF INVENTION: HUMAN EMRI-LIKE G PROTEIN COUPLED
TITLE OF INVENTION: RECEPTOR
FILE REFERENCE: PR-0550 US
CURRENT APPLICATION NUMBER: US/09/110,116
CURRENT FILING DATE: 1998-07-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 3
LENGTH: 886
TYPE: PRT
ORGANISM: HOMO SAPIENS
FEATURE:
OTHER INFORMATION: 784994, GenBank
US-09-110-116-3

```

	Query Match	11.1%	Score 222.5,	DB 3;	length 886;
	Best Local Similarity	27.9%:	Pred.	No.8.2e-10;	
	Matches	79;	Conservative	20;	Mismatches 95; Indels 89; Gaps 17;
OY	130 PGTVPDCLACGGGQRPSCSGNGHCSGDGSRGDSRCRHMGVQGGLCTCDMC--DGYS	186			
Dy	74 PGVRCKDIDCE-SQSPPQPGPNSSCK---NLISGRKXCSLDFSPSTGNWDPGAKRNPFS	129			
OY	187 SLRNETHSICFPCDESKTKTSGLTNRNDCGE-----EFGAVWLDEGAADVDD	232			
Dy	130 -----CIDDINE----C-LTSRVCPHEHSDCYNMSKGSYSCGVGFISRNSTCEDVN	174			

```

QY 233 ECAAEPPCSAAQCKCKNANSYTC-----ECCSSC---V 264
    ||:||||:|:|:|
Db 175 EC-ADPRACPEHATCNNTVGNYSFCENPFGESSGHLSCGLASCEDIDECEMCINS 233
    ||:||||:|:|:|
QY 265 GCTGEGPQN--CKECISGYAREHQ-----CADVDECSLAEKTVARKNENCYTPG 313
    ||:||||:|:|:|
Db 234 TCTNT-PSYTC-TCHPEFAPSSQLNFTDGVCECRDIDECRDPSTC-GPNSTCTNALG 290
    ||:||||:|:|:|
QY 314 SYVCVCPDGFEEETDACCVPAAEAETGESPTQ---LPSREDL 353
    ||:||||:|:|:|
Db 291 SYSCGCLVGFH-----PNEFSQKRGKNSCCQVLEKCKEDV 326
    ||:||||:|:|:|

```

```

1      RESULT 35
2      US-08-185-432-17
3      : Sequence 17, Application US/08185432
4      : Patent No. 5750552
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Artavanis-Tsakonas, Spyridon
9      : APPLICANT: Bussseau, Isabelle
10     : APPLICANT: Diederich, Robert J.
11     : APPLICANT: Xu, Tian
12     : APPLICANT: Matsuno, Kenji
13     : TITLE OF INVENTION: DEIFEX PROTEINS, NUCLEIC ACIDS, AND
14     : TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
15     : NUMBER OF SEQUENCES: 23
16     : CORRESPONDENCE ADDRESS:
17     : ADDRESSEE: PENNIE & EDMONDS
18     : STREET: 1155 Avenue of the Americas
19     : CITY: New York
20     : STATE: New York
21     : COUNTRY: U.S.A.
22     : ZIP: 10036-2711
23     :
24     : COMPUTER READABLE FORM:
25     : MEDIUM TYPE: Floppy disk
26     : COMPUTER: IBM PC compatible
27     : OPERATING SYSTEM: PC-DOS/MS-DOS
28     : SOFTWARE: PatentIn Release #1.0, Version #1.30
29     : CURRENT APPLICATION DATA:
30     : APPLICATION NUMBER: US/08/185,432
31     : FILING DATE: 21-JAN-1994
32     : CLASSIFICATION: 530
33     : ATTORNEY/AGENT INFORMATION:
34     : NAME: Mistrock, S. Leslie
35     : REGISTRATION NUMBER: 18,872
36     : REFERENCE/DOCKET NUMBER: 7326-006
37     : TELECOMMUNICATION INFORMATION:
38     : TELEPHONE: (212) 790-9090
39     : TELEFAX: (212) 869-8864/9741
40     : TELEFX: 66141 PENNIE
41     : INFORMATION FOR SEQ ID NO: 17:
42     : SEQUENCE CHARACTERISTICS:
43     : LENGTH: 2556 amino acids
44     : TYPE: amino acid
45     : TOPOLOGY: unknown
46     : MOLECULE TYPE: protein
47     :
48     : US-08-185-432-17

```

Query Match	11.1%	Score 222	DB 1:	Length 2556	
Best Local Similarity	24.6%	Pred. No. 2.9e-09			
Matches 75: Conservative	26	Mismatches 104	Indels 100	Gaps 15	
QY	125	KVCCSPGTY-GPDICACGGSGRPSGNHC-S-GGSRGGDSCRCRHMGVGLTDCMD	182		
	:	: : : : : :	:	:	:
Db	44	EACVCGGAFVGFRCDDPNCLSTPCKNAGTCHVDRGADYACSCALGFSGLPLTLPD	103		
	:	: : : : : :	:	:	:
QY	183	GYFSSLRNE-----THSICTACDESCKTSGLTNRDCG--CEYGMWLDEGACV-----	229		
	:	: : : : : :	:	:	:
Db	104	N-AALTNPGRNGRCDDLTLTRYRCRCPGMSGKSCQGDADPCASMPANGGCLPFEAS	161		
	:	: : : : : :	:	:	:
QY	230	-----DVDECAAEPPPCSAOFCNKANGSYTC-----	256		
	:	: : : : : :	:	:	:



Db 162 YICHCPSPFHGPTCQMDVNECGKPRLCRHGCTCHNEVGSYRCVCRATHHTGPCNCEPYPV 221  
QY 257 -----ECCDSSCV-GCTGEG-----PG-NCKE---CISGTARE 284  
Db 222 CSPSPCONGTCRPTGDTVTHEC--ACLPFTGQNCENIDDCPPNNCKNGACVGVNTY 279  
QY 285 HGOC-----ADVDECSLAECTVRKNENCTNPGSYVCVCPDGF-----EETEDA 329  
Db 280 NCPCPPEWTGQYCTEDYDECOLMPNAC-QNGCTCHNTHGTCVGVNGWGEDCSENID 338  
QY 330 CVPPA 334  
Db 339 CASAA 343

RESULT 36  
US-08-083-590A-20  
; Sequence 20, Application US/08083590A  
; Patent No. 5786158  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, S. et al.  
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
; TITLE OF INVENTION: Nucleic Acids  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/083,590A  
; FILING DATE: 25-JUN-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7326-015  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 8698864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2556 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-083-590A-20

Query Match 11.1%; Score 222; DB 1; Length 2556;  
Best Local Similarity 24.6%; Pred. No. 2.9e-09;  
Matches 75; Conservative 26; Mismatches 104; Indels 100; Gaps 15;  
QY 125 KVCSSPETY-GPDLACGGSQSPSCNGHCS-GDGSNQGSGCRCHMGYOGPLCTDCMD 182  
Db 44 EACVCGGAFVGPQRCDNPCLSTPCKNAGTCHVYDRGVADYACSCALGFSGPLCTPLD 103  
QY 183 GYFSSLNE-----THSICTADESCKTCSGLTNRDGE---CEVGVWLDGACAV----- 229  
Db 104 N--ACLTNPGRNGSTCDLTLTTEYKCKRCPGWSGSKCOQADPCANPCANGGCLPFEAS 161  
QY 230 -----DVDECAEPPEPCSAOFCCKNANGSYTC----- 256

Db 162 YICHCPSPFHGPTCQMDVNECGKPRLCRHGCTCHNEVGSYRCVCRATHHTGPCNCEPYPV 221  
QY 257 -----ECCDSSCV-GCTGEG-----PG-NCKE---CISGTARE 284  
Db 222 CSPSPCONGTCRPTGDTVTHEC--ACLPFTGQNCENIDDCPPNNCKNGACVGVNTY 279  
QY 285 HGOC-----ADVDECSLAECTVRKNENCTNPGSYVCVCPDGF-----EETEDA 329  
Db 280 NCPCPPEWTGQYCTEDYDECOLMPNAC-QNGCTCHNTHGTCVGVNGWGEDCSENID 338  
QY 330 CVPPA 334  
Db 339 CASAA 343

RESULT 37  
US-08-532-384-20  
; Sequence 20, Application US/08532384  
; Patent No. 6083904  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, S. et al.  
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And  
; TITLE OF INVENTION: Nucleic Acids  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/532,384  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/083,590  
; FILING DATE: 25-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7326-015  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 8698864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2556 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-532-384-20

Query Match 11.1%; Score 222; DB 3; Length 2556;  
Best Local Similarity 24.6%; Pred. No. 2.9e-09;  
Matches 75; Conservative 26; Mismatches 104; Indels 100; Gaps 15;  
QY 125 KVCSSPETY-GPDLACGGSQSPSCNGHCS-GDGSNQGSGCRCHMGYOGPLCTDCMD 182  
Db 44 EACVCGGAFVGPQRCDNPCLSTPCKNAGTCHVYDRGVADYACSCALGFSGPLCTPLD 103  
QY 183 GYFSSLNE-----THSICTADESCKTCSGLTNRDGE---CEVGVWLDGACAV----- 229

Db 104 N-ACLTNCRNGTCDLLTFLTEYKCRCPBGWSGKSCQADPCASNPANGGQLPEFAS 161  
 QY 230 -----DVDECALEPPPCSAOPCKNANGSYTC----- 256  
 Db 162 YICHCPSPFHGPTCRDVAEDVBCGKRLCRHGGCHNEVGSRYCVCRATHGPNCEPYPY 221  
 QY 257 -----EECDSCV-GCTGEG-----PG-NCKE---CISGYARE 284  
 Db 222 CSPSPQNGGTCTPRTDVTYHEC--ACLPFTGQNCENENIDDCPGNNCKNGACVGVNTY 279  
 QY 285 HGGC-----ADVDCSLAEKTCVRKNENCMTPGSYVCPDGF-----EETEDA 329  
 Db 280 NCPPEWGTGQCTEDVDYDEQLMPNAC-QNGGTCHNTHGGYCNVGNWGTGEDCSENID 338  
 QY 330 CVPFA 334  
 Db 339 CASAA 343

## RESULT 38

US-08-882-046-7  
 ; Sequence 7, Application US/08882046  
 ; Patent No. 6136952

## GENERAL INFORMATION:

APPLICANT: LI, Linheng  
 APPLICANT: Hood, Leroy  
 APPLICANT: Krantz, Ian D.  
 APPLICANT: Spinner, Nancy B.  
 TITLE OF INVENTION: Human Jagged Polypeptide, Encoding  
 TITLE OF INVENTION: Nucleic Acids and Methods of Use  
 NUMBER OF SEQUENCES: 110  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Campbell & Flores LLP  
 STREET: 4370 La Jolla Village Drive, Suite 700  
 CITY: San Diego  
 STATE: California  
 COUNTRY: USA  
 ZIP: 92122

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/882,046  
 FILING DATE: 25-JUN-1997  
 CLASSIFICATION: 536

## ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P-UW 2637  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 535-9001  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1010 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-882-046-7

Query Match 11.0%; Score 220.5; DB 4; Length 1010;  
 Best Local Similarity 22.9%; Pred. No. 1.4e-09;

Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 22;

QY 88 FECNQMLEAEHLEAMWLQKSEYDLP-EMFCVTLKVCSPGYGPDG--LAQGG 144  
 Db 194 FGCKNRCRPRD-----DFGCHYACDQNGNKTCEGMWGPCNNAICRQG- 237  
 QY 145 QRCPSGNHCSGDSRGDSCHHMGYQGPLCTDCM--DGYSSLRNE----- 191

Db 238 ---CSPK-----HGSCKLPDRCRCOYGMGLYCDKCIPIHPCGVHIGINEPWOCLBETNKG 289  
 QY 192 -----THSIC-----TACDE----- 201  
 Db 290 GOLCDKDLNVCYGHQPCNLGNGTCSNTGPKYQCSCEBGS-IGPNCIEAHEACLSDPCHNNG 349  
 QY 202 SCKTSGGLTNRDGCEGVWV-----LDE-----GAC----- 228  
 Db 350 SCKETS--LGFEFC-ECSPTWGTPTCTSTNIDDCSPNNCSHGTCQDLYNFKVCVCPQMTG 406  
 QY 229 ----VDVDCALAEPPPCSAOPCKNANGSYTC-----EECDSCVCTG--ECPGNC 274  
 Db 407 KTCQDLANECBAK--PCVNAKSKCNLIASYICDCLPGMNGQNCNDINIINDCLQCCQNDASG 464  
 QY 275 KE-----CISGYAREHGGCADVDECSLAERTCVKKNENCMTPGSYVCPDGF--- 323  
 Db 465 RDLVNGYRCICPPGYAGDHC- RDIDEC--ASNPCL-NGHQNONEINRQCCLPTGFSGN 520  
 QY 324 --EETEDACVP 332  
 Db 521 LCQLDIDYCEP 531

## RESULT 39

US-09-068-740A-6  
 ; Sequence 6, Application US/09068740A  
 ; Patent No. 6337387

## GENERAL INFORMATION:

APPLICANT: SAKANO, SEIJI  
 APPLICANT: ITOH, AKIRA  
 TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
 FILE REFERENCE: KP-8447  
 CURRENT APPLICATION NUMBER: US/09/068,740A  
 CURRENT FILING DATE: 1998-06-18  
 PRIOR APPLICATION NUMBER: JP 7-299611  
 PRIOR FILING DATE: 1995-11-17  
 PRIOR APPLICATION NUMBER: JP 7-311811  
 PRIOR FILING DATE: 1995-11-30  
 PRIOR APPLICATION NUMBER: PCT/JP96/03356  
 PRIOR FILING DATE: 1996-11-15  
 NUMBER OF SEQ ID NOS: 48  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 6  
 LENGTH: 1036  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-068-740A-6

Query Match 11.0%; Score 220.5; DB 4; Length 1036;  
 Best Local Similarity 22.9%; Pred. No. 1.4e-09;

Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 22;

QY 88 FECNQMLEAEHLEAMWLQKSEYDLP-EMFCVTLKVCSPGYGPDG--LAQGG 144  
 Db 163 FGCKNRCRPRD-----DFGCHYACDQNGNKTCEGMWGPCNNAICRQG- 206  
 QY 145 QRCPSGNHCSGDSRGDSCHHMGYQGPLCTDCM--DGYSSLRNE----- 191  
 Db 207 ---CSPK-----HGSCKLPDRCRCOYGMGLYCDKCIPIHPCGVHIGINEPWOCLBETNKG 258  
 QY 192 -----THSIC-----TACDE----- 201  
 Db 259 GOLCDKDLNVCYGHQPCNLGNGTCSNTGPKYQCSCEBGS-IGPNCIEAHEACLSDPCHNNG 318  
 QY 202 SCKTSGGLTNRDGCEGVWV-----LDE-----GAC----- 228  
 Db 319 SCKETS--LGFEFC-ECSPTWGTPTCTSTNIDDCSPNNCSHGTCQDLYNFKVCVCPQMTG 375  
 QY 229 ----VDVDCALAEPPPCSAOPCKNANGSYTC-----EECDSCVCTG--ECPGNC 274  
 Db 376 KTCQDLANECBAK--PCVNAKSKCNLIASYICDCLPGMNGQNCNDINIINDCLQCCQNDASG 433

```

Oy 275 KE-----CISYAREHGQCAVDSCSLAEKTCVRRKNEKCYTPGSSYCVCPDGE--- 323
      ::      |||  ::      |||  ::      |||  ::      |||  ||
Db 434 RDLVNGYRCICPPPIAGDHCE-RDIDEC--ASNFL-NGGHCCQNEINRFQCLPTGSGN 489
      ::      |||  ::      |||  ::      |||  ::      |||  ||
Oy 324 --EETEDACP 332
      ::      |||  ::      |||  ::      |||  ::      |||  ||
Db 490 LCQLDIDICEP 500

```

```

RESULT 40
US-09-068-740A-7
: Sequence 7, Application US/09068740A
: Patent No. 6337387
: GENERAL INFORMATION:
: APPLICANT: SAKANO, SEIJI
: APPLICANT: ITOH, AKIRA
: TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
: FILE REFERENCE: KP-8447
: CURRENT APPLICATION NUMBER: US/09/068, 740A
: CURRENT FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: JP 7-299611
: PRIOR FILING DATE: 1995-11-17
: PRIOR APPLICATION NUMBER: JP 7-311811
: PRIOR FILING DATE: 1995-11-30
: PRIOR APPLICATION NUMBER: PCT/JP96/03356
: PRIOR FILING DATE: 1996-11-15
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 7
: LENGTH: 1187
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-068-740A-7

```

Query Match	11.08	Score 220.5	DB 4	Length 1187
Best Local Similarity	22.98	Pred. No. 1.6e-09		
Matches 85	Conservative 33	Mismatches 94	Indels 159	Gaps 22

QY	88	FECONMTEAEOENLEAMWLLQKSEYRPLE-EMPCVKTLKVCSSPGTGYPRC--LAOOGGS	144
		:	:
Db	163	FGCNKFCPRP-----DFFGHTACDDQNGKNTCEGMWGPBECNRAICRQG-	206
QY	145	QRPSCGNHCSCGDSRGDSCRCMHGYOGLTDCM-DGYESSLNE-----	191
		:	:
Db	207	---CSPK-----HSCSKLRGDCRCROYEMQGLYCDKCLPHRCVCVNGICNEFMQCLCETNMG	258
QY	192	-----THSIC-----TACDE-----	201
		:	:
Db	259	GQLCDKDLNYCGTHQPCLINGTCSNTPDKYQSCSPREGSGSPNCEIAEHACLSDPCHNRG	318
QY	202	SCRKTCGLTINDCCECEYGV-----LDE-----GAC-----	228
		:	:
Db	319	SKETS--LGFEC-ECSPGWTGPRCSNIDDCSPNMSCSHGTCODLVNGFCYCPPMWTG	379
QY	229	---VDVECAEAPPPCSAAQFCFNANGSYNC-----EECOSCVGCTG--EGPGNC	274
		:	:
Db	376	KTCQLDANECAK--PCVNAKSCKNLLASTYCDCLPBGHMGONCDININDCLGQCQNDASC	433
QY	275	KE-----CISGYAREHGGACADVDECSLAETKCVKRNENCYNTPGSYCVCPDGF--	323
		:	:
Db	434	RDLVNGYRCICPPGYAGDHCE-RDIDEC--ASNPL-NGHCQCQEIINRFQCLCTPTGFSGN	489
QY	324	--EETEDACVP	332
		:	:
Db	490	LCQDLIDYCEP	500

RESULT 41  
US-08-400-159-6  
; Sequence 6, Application US/08400155  
; Patent No. 5869282  
; GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David  
 APPLICANT: Henrique, Domingos M. P.  
 APPLICANT: Lewis, Julian H.  
 APPLICANT: Myat, Anna M.  
 APPLICANT: Fleming, Robert J.  
 APPLICANT: Artavanis-Tsakonas, Spyridon  
 APPLICANT: Mann, Robert S.  
 APPLICANT: Gray, Grace E.  
 TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
 TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
 NUMBER OF SEQUENCES: 20  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/400.159  
 FILING DATE: 07-MAR-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 7326-029  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1218 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-400-159-6

Query Match	11.0%;	Score 220.5;	DB 2,	Length 1218;
Best Local Similarity	22.9%;	Pred. No. 1.7e-09;		
Matches 85; Conservative	33;	Mismatches 94;	Indels 159;	Gaps 22;

OY	88	FEONOMLEAOEHLNEMWLOLAKSEYPLDF-EMPEVKTLKVCSSGTYPGPD-	LAOCGGG	1444
	111:	---	---	---
Db	194	FGCNKFCPRD-	-BEFGHYACDDQGNKTCMBGMWPEPCNRAICRG-	237
OY	145	QRPSCGNHCSGSDSGRODGSRCRHMGQGPLCTIDCM--DGFSSLLNE	---	191
	11	111	111	111
Db	238	---CSPK-----HSCKRLPGDCRCQYHQGLYCDKCLTPHRCVCVAGICNEPQCLCETNMG	---	289
OY	192	---THSIC-----TACDE-----	---	201
	11	11	11	11
Db	290	GOLDCKDLNYCGTHQRPCLNGTGCSNTGPDYKQCSCPREGYSGNCEIEMHACLSDPCHNRG	---	349
OY	202	SCKTCGSLTLNRDGCCEYGVW----	LDE-----	GAC-----
	111	111	111	111
Db	350	SKETS--LGFEBC-ECSPGWTGPMPCSTNIDCSFNNGSHGGTCODLVNGFYKVCPPOMTG	---	228
OY	229	---VDVDECAAEPRPCSAAPFCNKNANSYTC-----	EDCDSVCVGTG--	EGPGNC
	11	11	11	11
Db	407	KTCOLDANECEAK--PCVNNAASCKNLLASTYCDCLPGMGQNCNDININDLGGQCQNDASC	---	466
OY	275	KE-----CISGAREHGCACVADDECSLAEKTCVKNRKNICYMTPGSYVCVCPDGF---	---	323
	11	111	111	111
Db	465	RDLVNGYACICLPPGAGDHCE-RIDDEC--ASNPL--NGHCQNEINRFQCLCPTGFSGN	---	520
OY	324	---EETEDACVP	---	332

Db 521 LCQLDIDYCEP 531

RESULT 42

US-08-611-729A-6  
; Sequence 6, Application US/08611729A  
; Patent No. 6004924

GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
APPLICANT: Gray, Grace E.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,729A  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-037  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 866-9741/8864  
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1218 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-611-729A-6

Query Match 11.0%; Score 220.5; DB 3; Length 1218;

Best Local Similarity 22.9%; Pred. No. 1.7e-09;  
Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 22;

Db 88 FPCNOMLEAEOEHLBAMWLQKSEYDLE-EMFVCVTKLVCCSPGTYPDC--IACQGS 144

194 FCGNFCRRRD-----DFGHTACDQNGNKTCEGMGPECNRAICRQG- 237

QY 145 QPCSGNGHCSGDGRGDSRCRHMGYGPLCTDCM--DGYFSSLRNE----- 191

238 ---CSPK---HGSKCLPGDCRCQYGMQGLYCDKICIPHPCGVHIGICNPMQCLCETNMG 289

QY 192 -----THSIC-----TACDE----- 201

Db 290 GOLCDKDLNYGCTHQPCLNGTCSNTGPDKYQCSCEBGYSGPNCETIAEHACLSDPCHNRG 349

QY 202 SKTCSGLTNRDGCCEVGMV-----LDE-----GAC----- 228

Db 350 SKETS--LGFEC-ECSPTGPTGTCSTNIDDCSPNNCSHGCTICDQLVNGFKVCVPPQWTG 406

QY 229 ---VDVDECAEPPECASAOFCNNANGSYTC-----EECDSSVCGCTG--EGPNC 274

Db 407 KTCQDLANDCEAK--PCVNAKSKNLIASYCDDLPGWMGQVCDININDICGQONDSAC 464

QY 275 KE-----CISGTAREHGQCADVDECSLAETKTVRKNNKNNYINPGSYVCCPDGF--- 323

Db 465 RDLVNGYRCICPPGYAGDHCE-RDIDEC--ASNPL-NGGHQNEINRPOCLCPTGFSGN 520

QY 324 --EETEDACVP 332

Db 521 LCQLDIDYCEP 531

RESULT 43

US-08-882-046-2  
; Sequence 2, Application US/08882046  
; Patent No. 6136952

GENERAL INFORMATION:

APPLICANT: Li, Linheng  
APPLICANT: Hood, Leroy  
APPLICANT: Krantz, Ian D.  
APPLICANT: Spinner, Nancy B.  
TITLE OF INVENTION: Human Jagged polypeptide, Encoding  
TITLE OF INVENTION: Nucleic Acids and Methods of Use  
NUMBER OF SEQUENCES: 110  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/882,046  
FILING DATE: 25-JUN-1997  
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-UW 2637  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1218 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-882-046-2

Query Match 11.0%; Score 220.5; DB 4; Length 1218;

Best Local Similarity 22.9%; Pred. No. 1.7e-09;  
Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 22;

Db 88 FPCNOMLEAEOEHLBAMWLQKSEYDLE-EMFVCVTKLVCCSPGTYPDC--IACQGS 144

194 FCGNFCRRRD-----DFGHTACDQNGNKTCEGMGPECNRAICRQG- 237

QY 145 QPCSGNGHCSGDGRGDSRCRHMGYGPLCTDCM--DGYFSSLRNE----- 191

Db 238 ---CSPK---HGSKCLPGDCRCQYGMQGLYCDKICIPHPCGVHIGICNPMQCLCETNMG 289

QY 192 -----THSIC-----TACDE----- 201

Db 290 GOLCDKDLNYGCTHQPCLNGTCSNTGPDKYQCSCEBGYSGPNCETIAEHACLSDPCHNRG 349

Oy	202	SCKFCSSGLTINDDCECEFGW	-----LDE	-----GAC	-----	228
Db	350	SKETLS--LGEFC--ECSEBGMTPTCTSTNIDDCSPNNCSHGTCODLVNGFKCYCPQWMTG				406
Oy	229	-----VDVDECAEAPPSPSAOFCCKNANGSYTC	-----	EECDSSCVGCTG--	EGPGNC	274
Db	407	KTCQDLANECSAK--PCVMAKSCKNLLASTAYCCDCLPGPMQONCEDININDCLGQQQNDASC				464
Oy	275	KE-----CISGYAREHGOCADVDEBCSLAECTVARKNENCYNTPGSYVCVCPDGF--				323
		:::	:::	:::	:::	
Db	465	RDLVNGYKCLCPPYAGDHCE--RDIDEC--ASNPL--NGHCQNEIRNFQCLPTGSGN				520
Oy	324	-----EETEDACVP	332			
Db	521	LCQLDIDYCEP	531			

```

: RESULT 44 278-7
: US-09-214-278-7
: Sequence 7, Application US/09214278
: Patent No. 6291210
: GENERAL INFORMATION:
: APPLICANT: Sakano, Seiji
: APPLICANT: Itoh, Akiya
: TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
: FILE REFERENCE: KP-8576
: CURRENT APPLICATION NUMBER: US/09/214,278
: CURRENT FILING DATE: 1999-01-26
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO: 7
: LENGTH: 1218
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-214-278-7

```

Query Match	Similarity	11.0%	Score	220.5	DB	4	Length	1218
Best Local	Similarity	22.9%	Pred. No.	1.7e-09				
Matches	85	Conservative	33	Mismatches	94	Indels	159	Gaps

  

QY	88	EFCNMLEAQEHLEAMWLQKSEYRDLF-EMFCVKTLKVVCCSPGYGPDG--LACGGGS	144
		:::	
Db	194	FGCNMFCRPRD-----DFGHIACDQNGKTKMEGMGPECNRNATCRQG	237
		:::	
QY	145	QRPCGNGHSCDSDSGROGDSCRCHMGYGRPLCTDCM--DGYFSSLRNE-----	191
		:::	
Db	238	---CSPK-----HGCKLPGDRCQYGMWGLYCDKCIAPRGCVHIGCNEPMOCLERTNWG	289
		:::	
QY	192	-----THSIC-----TACDE-----	201
		:::	
Db	290	GOLDCKDLNYGCTGHPCLNGGTCMTGPRDKYQCSCEPGLSGNSEIAEHACLSDPCHNNG	349
		:::	
QY	202	SCKTCGSLTNRDGCEVGWV-----LDE-----GAC-----	228
		:::	
Db	350	SKERTS--LGFEPC-BCSPGWTGPTCSFNTNIDDCSPNNCSHGTCODLVNFKVCYCPQWNG	406
		:::	
QY	229	---VDVDECAEERPSCSAORCKNANSGSYTC-----EECDSSCVGCTG--BEPPNG	274
		:::	
Db	407	KTCQLDANECEK--PCVNAKSKMLIASYDCCLPGRMGWQCKDITINDCLDQCCQADASC	464
		:::	
QY	275	KE-----CISGYAREHGQCADVDEGCSLAERTCVRKNEKNCYNTPGSYVCVCPDGF--	323
		:::	
Db	465	RDLVNGYRCICPPGYAGDHCE-RDIDEC--ASNPCI-NGCHQONEINRQCLCPYFSFN	520
		:::	
QY	324	---EETEDACVP	332
		:::	
Db	521	LCQLDIDYCEP	531

```

: Patent No. 6337387
: GENERAL INFORMATION:
: APPLICANT: SAKANO, SEIJI
: APPLICANT: ITOH, AKIRA
: TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
: FILE REFERENCE: KP-8447
: CURRENT APPLICATION NUMBER: US/09/068,740A
: CURRENT FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: JP 7-299611
: PRIOR FILING DATE: 1995-11-17
: PRIOR APPLICATION NUMBER: JP 7-311811
: PRIOR FILING DATE: 1995-11-30
: PRIOR APPLICATION NUMBER: PCT/JP96/03356
: PRIOR FILING DATE: 1996-11-15
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: Patentl Ver. 2.1
: SEQ ID NO 11
: LENGTH: 1218
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-068-740A-11

```

Query Match	11.0%;	Score 220.5;	DB 4;	Length 1218;
Best Local Similarity	22.9%;	Pred. No. 1,7e-09;		
Matches	85;	Conservative	33;	Mismatches 94; Indels 159; Gaps 22

  

Oy	88	FEENOMLEAQEHLEAMWLQKSEYRDLF-EMFCVYTKLVCCSPGTGPGDC--LACQGGGS	144
		:	:
Db	194	FGCNCFRPRD-----DEFGHYACDQNGKTKTEMGWGEPCNRATCRQ-	237
		:	:
Oy	145	QRPGSGNGHCSGDGSGRQDGSRCRHMGVQGPLCTDCM--DGYFSSLRN-	191
		:	:
Db	238	---GSPK-----HGSCKRLPGDRCQYGMWGLYCDKCIPIRPGCVHIGICNEPPOCLCETNMG	289
		:	:
Oy	192	-----THSIC-----TACDE-----	201
		:	:
Db	290	GOLCDKDLNYCCTHQPLCLNGGTCNTMGPRDKYQCSPEGYSGPNCIEAHACLSDPCHNNG	349
Oy	202	SKCTSGSLTNRCGCEVGVN-----LDE-----GAC-----	228
		:	:
Db	330	SKETNS--LGFEPC-BCSPMTGTPTCSNTDDCSPNHCSHGFCODLVNMGFKVCVCPROWTG	406
Oy	229	---VDVDECAAEPPCSAOFCKNANGSYTC-----EECDSSVCVCTG--EAGPNC	274
		:	:
Db	407	KTCQDANAECEAK--PCVNAKSCKNULIASIYYDCLPGMMGQMDININDLQOQNDASC	464
Oy	275	KE-----CISGYAREHGGQCADVDECSLAEKTCVRKNENCVNTPGSYVCVCPDGF---	323
		:	:
Db	465	RDLVNGYRCICPPGYAGDHCE--RDIDEC--ASNPTCL-NGGHQCNETINRCQCLCPGFGSGN	520
Oy	324	---EETEDACVP	332
		:	:
Db	521	LCQDLIDYCEP	531

```

RESULT 46
5258288-1
PATENT No. 5258288
APPLICANT: Wydro Robert; Cohen Edward; Dackowski William
Stenflo John; Lundwall Ake; Dahlback Bjorn
TITLE OF INVENTION: VECTOR CONTAINING DNA ENCODING MATURIN
HUMAN PROTEIN S
NUMBER OF SEQUENCES: 15
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/890,401
FILING DATE: 25-JUL-1986
SEQ ID NO: 1
LENGTH: 675
5258288-1

```

RESULT 45  
US-09-068-740A-11  
; Sequence 11, Application US/09068740A

Query Match 10.98; Score 218.5; DB 6; Length 675;

Best Local Similarity 28.8%; Pred. No. 1.3e-09;  
Matches 69; Conservative 23; Mismatches 73; Indels 75; Gaps 16;

QY 112 YPLPFMPFCYKTLKACGSPGTYGPDCLACGGSGGSGNGHCS-DGSGSGDSCGRCH 170  
DB 107 YPLPRLS-CVNAISDQCNP-----LPCNEDGFMTCKDG-QATFTCLGCS 147  
QY 171 GYOGPLC---TDMDGYFSSLRNETHSICTACDESKTCSGLTNRDCEGEVWVL-D 224  
DB 148 GMGEGCESDINCKDPV-----NINGGCSQJCEMTPG--SYHC-SCKNGFVWLSN 195  
QY 225 EGCACVDECAAPPCSAAGFCRNANGSTCECDSGCVGCTGEGPKNKECTISYARE 284  
DB 196 KPOCKQVDECVLRPSICGTA-VCKNIPGDEEC-----ECAEGKYXN 235  
QY 285 --HGCAADVDECSLAETCYCRKNENCYNTPGSYVCV--PDGFEETED-----ACVP 332  
DB 236 PVKSCDDVDEEC-AENMLCA--QLCVNTPGSGTCTCDGKKFKLADQKSCAIVYCLP 290

## RESULT 47

US-08-185-432-19  
Sequence 19, Application US/08185432  
Patent No. 5750652

## GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Bussieu, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530

## ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2703 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-19

Query Match 10.8%; Score 216.5; DB 1; Length 2703;  
Best Local Similarity 23.3%; Pred. No. 8.3e-09;  
Matches 97; Conservative 27; Mismatches 135; Indels 157; Gaps 20;

QY 41 FNGQAMVDTAKN-----FGGNTAMEEKTLSKY-----ESSEIR 74  
DB 130 FDESLEIAVPNACDHVTCINGTC--QLKTLBYTCACANGYTGERCTKNICLASSPCR 187

QY 75 LLEILEGCESSDPECN-----QMLEAOEHLIEMWMLKSEYPLDFEMFCVK 122  
DB 188 NGATCTALAGSSSEFTSCPPGFTGDTCTSYDIEBCSN-----PKYGGICVN 234  
QY 123 T---LKVCCSPGTYGPDG---LACGGSGORPCSGNGHCSGDSKRGDSCCHMGYOP 175  
DB 235 THGSYOCMEPTGYTGKDCCTKYPK--SPSPQONAGTGRSNGLSY--ECKCPKGFEGK 288  
QY 176 LCTDMDGYFSSL-RNETHSICTACDESKTCSGLTNRDCEGEVWVLDEGACVVDVDC 234  
DB 289 NCEQNTDDCLGHLCQNGGTCIDGISYTCRCPPNFTGRCCD-----DYDVC 335  
QY 235 A-AEPPCSAAGFCRNANGSTCECDSGCV-GCTG---EEPGNKE-----CISG 280  
DB 336 AQRDHPVCGNGATCTNTHGSYC-----ICVNGMAGLDCSNMTDCKQAACFYATCIDG 390  
QY 281 YAREHQCA-----VY 291  
DB 391 VGSFYCQCTKTKTGLLCHLDACSNPCHADACTIDTSPINGSYACSCANGYGVDCSEDI 450  
QY 292 DECSLAETCYCRKNENCYNTPGSYVCVCPDGFEETEDACVPPAEATGES-PTQ 346  
DB 451 DECD--QGSPECHNGICVNTPGSYRCNCSQGFITG-----PRCEININCESHPCQ 498

## RESULT 48

US-08-185-432-16  
Sequence 16, Application US/08185432  
Patent No. 5750652

## GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Bussieu, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530

## ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-16

Query Match 10.7%; Score 215.5; DB 1; Length 2471;

Best Local Similarly 21.8%; Pred. No. 9e-09;  
Matches 88; Conservative 43; Mismatches 110; Indels 163; Gaps 23;

```
OY 28 PTPCH--RCRLGVKFNQGMVDTAKKNGGNTAMEETLSKYES---SEIRLEIIEGL 82
    |||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 576 PDPCHHGQCQGDIDSY-----TCICNPGYGAICSDQIDECYSSPCLNDRCIDLVNG- 628
OY 83 CESSDFECNOMLEAQEEHLEAMWLQLKSEYPDLEMFVCVTKLVKCCSPGTGPDC-LACQ 141
    ::|||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 629 -----YQCN-----CQGTSGVNCIEINFD 647
OY 142 GGSORPCSGNGHSGDGRQDGSRCRMGYQGPLC-----TDCMDGYFS 186
    :|||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 648 DCASNPCI-HGICMDGIRY---SCVCSPGFTGRCNIDIDECASNPCRKATCINGV-- 701
OY 187 SLRNETHSIC-----TACDESKTC-----SGLNRRCGBEVGVW----LD 224
    |||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 702 ---NGFRICPEGRPHHPSQYVNECLSNPCIHGNCGTGLSGYKC-LCDAGWGINCEVD 757
OY 225 EGAC-----VDVDECAAEPPCSAAQFCCKNANG 252
    :|||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 758 KNECLSNPCQNGTCDNLVNGYRCTCKKGFKGYNQVNIIDECASN--PCLNQGTCFDDIS 815
OY 253 SYTC-----EBCDSSCVGCTG---EGPNCKE-----CISGYAREHGQA-DVD 292
    |||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 816 GYTCVLPYTGKNCQYVLAPCSNPCEANAIVCKESPFESEYTCICAPGMQGORCTIDID 875
OY 293 ECSLAETCVRKNECYNTPGSYVCVCPDGF-----EETEDACV 331
    |||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 876 EC--ISKPCM-NHGLCHNTQGSYMCBPCPGSGMDCEEDIDCL 916
```

RESULT 49  
US-08-083-590A-19  
Sequence 19, Application US/08083590A  
Patent No. 5786158

## GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie &amp; Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/083,590A

FILING DATE: 25-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 8698864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 2471 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-083-590A-19

## Query Match

10.7%; Score 215.5; DB 1; Length 2471;

Best Local Similarly 21.8%; Pred. No. 9e-09;

Matches 88; Conservative 43; Mismatches 110; Indels 163; Gaps 23;

```
OY 28 PTPCH--RCRLGVKFNQGMVDTAKKNGGNTAMEETLSKYES---SEIRLEIIEGL 82
    |||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 576 PDPCHHGQCQGDIDSY-----TCICNPGYGAICSDQIDECYSSPCLNDRCIDLVNG- 628
OY 83 CESSDFECNOMLEAQEEHLEAMWLQLKSEYPDLEMFVCVTKLVKCCSPGTGPDC-LACQ 141
    ::|||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 629 -----YQCN-----CQGTSGVNCIEINFD 647
OY 142 GGSORPCSGNGHSGDGRQDGSRCRMGYQGPLC-----TDCMDGYFS 186
    :|||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 648 DCASNPCI-HGICMDGIRY---SCVCSPGFTGRCNIDIDECASNPCRKATCINGV-- 701
OY 187 SLRNETHSIC-----TACDESKTC-----SGLNRRCGBEVGVW----LD 224
    |||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 702 ---NGFRICPEGRPHHPSQYVNECLSNPCIHGNCGTGLSGYKC-LCDAGWGINCEVD 757
OY 225 EGAC-----VDVDECAAEPPCSAAQFCCKNANG 252
    :|||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 758 KNECLSNPCQNGTCDNLVNGYRCTCKKGFKGYNQVNIIDECASN--PCLNQGTCFDDIS 815
OY 253 SYTC-----EBCDSSCVGCTG---EGPNCKE-----CISGYAREHGQA-DVD 292
    |||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 816 GYTCVLPYTGKNCQYVLAPCSNPCEANAIVCKESPFESEYTCICAPGMQGORCTIDID 875
OY 293 ECSLAETCVRKNECYNTPGSYVCVCPDGF-----EETEDACV 331
    |||  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 876 EC--ISKPCM-NHGLCHNTQGSYMCBPCPGSGMDCEEDIDCL 916
```

RESULT 50  
US-08-532-384-19  
Sequence 19, Application US/08532384  
Patent No. 6083904

## GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie &amp; Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/532,384

FILING DATE: 25-JUN-1993

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 8698864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-532-384-19

Query Match 10.7%; Score 215.5; DB 3; Length 2471;  
Best Local Similarity 21.8%; Pred. No. 9e-09;  
Matches 88; Conservative 43; Mismatches 110; Indels 163; Gaps 23;

QY 28 PIPCH--RCRGVDRKFNQGVNVDPAKKNPFGGNTAMWEKILSKYES--SEIRLLEILEGI 82  
DB 576 PDCPHHGQCDQDIDSY-----TCICNPGYMAICSDQIDECTSSPCLNDGRICIDLNG- 628  
QY 83 CESSDEECNOMLEAQEEHLEAMWLQKSEYDLEFWFCVKTLLKCCSPGTGYPDC-LACQ 141  
DB 629 -----YQCN-----CQPGTSGVNCIINF 647  
QY 142 GGSQRPCSGNGHCSGDSRGDSGRCHMGYQGPLC-----TDAMDGYFS 186  
DB 648 DCASNPCI-HGICMDGINRY---SCVCSPGFTGRCNIDIDECASNPCRKATCINGV-- 701  
QY 187 SLRNETHSIC-----TACDESCKTC-----SGLTNRDCGCEYGVV-----LD 224  
DB 702 ---NGFRCLCPPEGPHHPSCISOVNECLSNPCINGNCTGGLSGYKC-LCDAGWVGINCEVD 757  
QY 225 EGAC-----VDVDECAAEPPPCSAQFCRNANG 252  
DB 758 KNECLSNPCQNGGTCDNLVNGYRCTCKKGFKNQCVNIDECASN--PCLNGTCTCFDDIS 815  
QY 253 SYTC-----ECCDSSCVGCTG---EGRGNKE-----CISGYAREHQCA-DVD 292  
DB 816 GYTCHCVLPYTGKNCQTVLAPCSPNPECENAAVCKESPNEFSYTCICAPGWGGQRCITDID 875  
QY 293 ECSIAERTCVKKNENYNFPGSYVYVCPDGF-----EETEDACV 331  
DB 876 EC--ISKPCM-NHGLCHNTGSGTYMCECPFGSGMDCEEDIDCL 916

Search completed: September 10, 2002, 11:08:46  
Job time: 588 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 11:01:38 ; Search time 32.64 Seconds  
(without alignments)  
1201.258 Million cell updates/sec

Title: US-09-905-075-2  
Perfect score: 2005  
Sequence: 1 MRLPRRAAGLPLLLLP.....AEAEATGEGSPQLPSREDL 353

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 65 summaries

Database :  
A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2005	100.0	353 20	AAV08064 Human EGF-like hom
2	2005	100.0	353 20	AAV13344 Amino acid sequenc
3	2005	100.0	353 20	AAV05283 EGF-like homologue
4	2005	100.0	353 21	AAV05169 PRO211 polypeptide
5	2005	100.0	353 21	AAV83224 PRO211 polypeptide
6	2005	100.0	353 22	AAV12319 Human PRO211 polyp
7	2005	100.0	353 22	AAV8596 PRO211 Homo sapi
8	2005	100.0	353 22	AAV80212 Human PRO211 prote
9	2005	100.0	353 22	AAV81231 Human TANGO 31 pr
10	2005	100.0	353 22	AAV53075 Human angiotensin
11	1997	99.6	353 21	AAV88571 Human PRO211 amino

12	1887	94.1	329	22	AAV61233	Mature human TANGO
13	1716	85.6	318	21	AAV42711	Human OREF ORF2475
14	1647	82.1	353	21	AAV91870	Human apoptosis re
15	1573	78.5	386	22	AAV1870	Novel human secret
16	1481.5	73.9	348	22	AAV61240	Chinese hamster pr
17	972.5	48.5	420	22	AAV8106	Human TANGO 206 po
18	971.5	48.5	420	20	AAV13362	Amino acid sequenc
19	971.5	48.5	420	20	AAV05281	EGF-like homologue
20	971.5	48.5	420	21	AAV82396	Human PRO214 prote
21	971.5	48.5	420	21	AAV88569	Human PRO214 amino
22	971.5	48.5	420	22	AAV59899	Human polypeptide
23	971.5	48.5	420	22	AAV12316	Human PRO214 polyp
24	971.5	48.5	420	22	AAV68594	PRO214. Homo sapi
25	971.5	48.5	420	22	AAV80230	Human PRO214 prote
26	971.5	48.5	420	22	AAV84813	Human TANGO 206 va
27	971.5	48.5	420	22	AAV87228	Human EXMAD-6 SEQ
28	971.5	48.5	513	22	AAV41685	Human polypeptide
29	969.5	48.4	420	22	AAV84814	Human TANGO 206 va
30	969.5	48.4	420	22	AAV84815	Human TANGO 206 va
31	969	48.3	434	21	AAV83835	Human secreted pro
32	969	48.3	434	21	AAV76151	Human secreted pro
33	968.5	48.3	420	22	AAV88110	Mouse TANGO 206 po
34	967.5	48.3	420	22	AAV84817	Mouse TANGO 206 va
35	966.5	48.2	420	22	AAV84816	Mouse TANGO 206 va
36	965.5	48.2	420	22	AAV84818	Mouse TANGO 206 va
37	860.5	42.9	392	21	AAV83839	Human secreted pro
38	670	33.4	374	21	AAV85688	Drosophila melanog
39	463.5	23.1	242	21	AAV76114	Rat HT glycoprotein
40	463.5	23.1	242	21	AAV56053	Skin cell protein
41	406	20.2	140	22	AAV811766	Human apoptosis re
42	274.5	13.7	2189	11	AAV05222	Antigen GX5401FL e
43	264.5	13.3	637	21	AAV32346	Mouse cell surface
44	266.5	13.3	644	21	AAV79186	Haematopoietic ste
45	266.5	13.3	644	21	AAV79193	Haematopoietic ste
46	260	13.0	1121	21	AAV78887	Human fibrillin 3
47	256.5	12.8	1208	21	AAV70554	Human latent trans
48	256.5	12.8	1257	21	AAV70554	Human latent trans
49	256.5	12.8	1260	22	AAV81460	TANGO 275 mature p
50	256.5	12.8	1289	22	AAV81419	Human TANGO 275 p
51	255.5	12.7	1253	21	AAV12271	Mouse latent TGF- $\beta$
52	255.5	12.7	1253	21	AAV12271	Murine TANGO 275 p
53	254.5	12.7	1256	18	AAV61420	Human fibrillin type
54	254.5	12.7	566	12	AAV11148	Fibrillin A. Homo s
55	254.5	12.7	601	12	AAV11149	Fibrillin B. Homo s
56	254.5	12.7	601	18	AAV27599	Human fibrillin type
57	254.5	12.7	652	19	AAV49879	Amino acid sequenc
58	254.5	12.7	652	21	AAV32345	Human cell surface
59	254.5	12.7	683	18	AAV11150	Fibrillin C. Homo s
60	254.5	12.7	683	18	AAV27600	Human fibrillin type
61	254.5	12.7	686	22	AAV3385	Novel human diagn
62	254.5	12.7	703	18	AAV27601	Human fibrillin type
63	254.5	12.7	706	22	AAV3386	Novel human diagn
64	253	12.6	3680	22	AAV70878	Drosophila melanog
65	250	12.5	1833	16	AAV79478	Mouse LTRP-2. Mus

## ALIGNMENTS

RESULT 1  
ID AAV08064 standard; Protein: 353 AA.  
AAV08064:  
11-SEP-2000 (first entry)  
Human EGF-like homologue protein (PRO217) encoded by DNA32292 cDNA.  
Inflammatory cell infiltration; immune response; T cell proliferation;  
anti-inflammatory; anti-autoimmune; anti-diabetic; spondyloarthritis;  
T cell-mediated disease; spondyloarthritis; sclerodermis; renal disease;  
Inflammatory myopathy; hemolytic anemia; thrombocytopenia; thyroiditis;  
KW

KW diabetes mellitus; demyelinating polyneuropathy; Guillain-Barre syndrome;  
 KW multiple sclerosis; polyneuropathy; hepatitis; cirrhosis; enteropathy;  
 KW sclerosing cholangitis; inflammatory bowel disease; Whipple's disease;  
 KW skin disease; dermatitis; psoriasis; asthma; allergic rhinitis; tumor;  
 KW food hypersensitivity; urticaria; eosinophilic pneumonia; transplant;  
 KW idiopathic pulmonary fibrosis; graft rejection; PRO245; human; PRO217;  
 KW EGF-like.  
 OS Homo sapiens.  
 XX MO9914241-A2.  
 PN 25-MAR-1999.  
 PD 17-SEP-1998; 98WO-US19437.  
 XX PF 17-SEP-1997; 97US-0059119.  
 PR 18-SEP-1997; 97US-0059263.  
 PR 28-OCT-1997; 97US-0063350.  
 PR 12-NOV-1997; 97US-0065186.  
 PR 21-NOV-1997; 97US-0066364.  
 PR 24-NOV-1997; 97US-0066770.  
 PR 04-JUN-1998; 98US-0088026.  
 XX (GETH ) GENENTECH INC.  
 PA Fong S, Goddard A, Gurney AL, Tumas D, Wood WI;  
 PI WPI: 1999-2294499/19.  
 DR N-PSDB; MAX37671.  
 XX Composition containing novel polypeptide PRO245, its agonist or  
 PT antagonist -  
 PT Example 1; Fig 6B; 177pp; English.  
 PS This invention describes a novel composition containing (part from a  
 CC carrier or excipient), a novel PRO245 polypeptide (I), its agonist or  
 CC antagonist, or their fragments, for modulating: (i) infiltration of  
 CC inflammatory cells into tissue; (ii) an immune response; or (iii) T cell  
 CC proliferation. The composition increases or decreases any of the effects  
 CC (i)-(iii). The products of the invention have anti-inflammatory,  
 CC anti-autoimmune and anti-diabetic activity. (I), and its (ant)agonists  
 CC and their fragments, are used to treat immune-related diseases,  
 CC particularly T cell-mediated diseases. The diseases treated include  
 CC systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic  
 CC arthritis, spondyloarthropathies, systemic sclerosis (scleroderma),  
 CC idiopathic inflammatory myopathies (dermatomyositis, polymyositis),  
 CC Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune  
 CC hemolytic anemia (immune pancytopenia, paroxysmal nocturnal  
 CC hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic  
 CC purpura immune-mediated thrombocytopenia), thyroiditis (Grave's disease,  
 CC Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic  
 CC thyroiditis), diabetes mellitus, immune-mediated renal disease  
 CC (glomerulonephritis, tubulointerstitial nephritis), multiple sclerosis,  
 CC idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome, chronic  
 CC inflammatory demyelinating polyneuropathy, infectious hepatitis  
 CC (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune  
 CC chronic active hepatitis, primary biliary cirrhosis, granulomatous  
 CC hepatitis, and sclerosing cholangitis. Inflammatory bowel disease  
 CC (ulcerative colitis; Crohn's disease), gluten-sensitive enteropathy, and  
 CC Whipple's disease. Autoimmune or immune-mediated skin diseases including  
 CC bullous skin diseases, erythema multiforme, contact dermatitis, psoriasis,  
 CC asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity,  
 CC urticaria, eosinophilic pneumonia, idiopathic pulmonary fibrosis,  
 CC hypersensitivity pneumonitis, and transplantation associated diseases  
 CC (graft rejection, and graft-versus-host-disease). (I), its (ant)agonists  
 CC or fragment can also be used as an adjuvant in treatment of tumors.  
 CC Antibodies against (I) can also be used for diagnosing such diseases.  
 CC This sequence represents a human EGF-like homologue (PRO217) encoded by  
 CC cDNA clone DNA2292 which is described in the invention.  
 XX Sequence 353 AA:

Query Match 100.0%; Score 2005; DB 20; Length 353;  
 Best local similarity 100.0%; Pred. No. 1.8e-133;  
 Matches 353; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRLPRRALGLPLLLLPAPBAKAPPPCHRCGLVDKFMQGVDTAKKNGGNTAM 60  
 DB 1 mrlpraalglpllllpapbaakppchrcrglvdkfmqgvdtakkngfgntaw 60  
 QY 61 EEKTLISKYSSSEIRLEILEIGCESSDFECNQMLEAOEHELEAMVLTOLKSEYPDLEFEMRC 120  
 DB 61 eektliskyessseirllleileigleessdfecnqmlaeqheelawltqlkseypdllefrc 120  
 QY 121 VKTLKVCSPGTYGPDCLACQGGSPRCGNGHCSGDSRQGGSCRCMHYQGPPLCTDCC 180  
 DB 121 vktlkvcspgtygpdclacqggsqpcsgnngchsgdsrgqdgscrcmhyygqplctdc 180  
 QY 181 MDGYFSSLLNETHSICTACDESKTCSGLTNRDGCSEYGVWLDGACVYDDECAAEPPP 240  
 DB 181 mdgyfssllnethsictacdescktsqgltnrdgcsevgwldgacvdydecaaeppp 240  
 QY 241 CSAAOCCKNANGSYNCEEDSSCVGCTGSGPGMKCKECISGYAREHGOCADVDPCSAEKT 300  
 DB 241 csaaoccknangsyntceedsscvgctgsgpgmkckecisgyarehgcadvdpcsaekt 300  
 QY 301 CVRKNNCYNTPGSYVCVCPDGFEEFEDACVPPAEAEATGESPTQLPSREDL 353  
 DB 301 cvrkncncyntpgsyvcvcpdgdfeetedacvppaeaeategesptqlpredl 353  
 RESULT 2  
 AAY13344  
 ID AAY13344 standard; Protein: 353 AA.  
 AC AAY13344:  
 DT 25-JUN-1999 (first entry)  
 XX Amino acid sequence of protein PRO211.  
 KW Secreted protein; transmembrane protein; human; enterocolitis;  
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;  
 KW congenital microvillus atrophy; skin disease; cell growth;  
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophila areata;  
 KW anti-thrombotic; wound healing; tissue repair.  
 OS Homo sapiens.  
 XX MO9914328-A2.  
 PN 25-MAR-1999.  
 PD 16-SEP-1998; 98WO-US19330.  
 XX PF 25-NOV-1997; 97US-0066840.  
 PR 17-SEP-1997; 97US-0059113.  
 PR 17-SEP-1997; 97US-0059115.  
 PR 17-SEP-1997; 97US-0059117.  
 PR 17-SEP-1997; 97US-0059119.  
 PR 17-SEP-1997; 97US-0059121.  
 PR 17-SEP-1997; 97US-0059122.  
 PR 17-SEP-1997; 97US-0059184.  
 PR 18-SEP-1997; 97US-0059263.  
 PR 18-SEP-1997; 97US-0059266.  
 PR 15-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 17-OCT-1997; 97US-0062287.  
 PR 24-OCT-1997; 97US-0063486.  
 PR 24-OCT-1997; 97US-0062814.  
 PR 24-OCT-1997; 97US-0062816.

PR 24-OCT-1997; 97US-0063045.  
 PR 24-OCT-1997; 97US-0063120.  
 PR 24-OCT-1997; 97US-0063121.  
 PR 24-OCT-1997; 97US-0063127.  
 PR 24-OCT-1997; 97US-0063128.  
 PR 27-OCT-1997; 97US-0063329.  
 PR 27-OCT-1997; 97US-0063327.  
 PR 28-OCT-1997; 97US-0063341.  
 PR 28-OCT-1997; 97US-0063342.  
 PR 28-OCT-1997; 97US-0063344.  
 PR 28-OCT-1997; 97US-0063349.  
 PR 28-OCT-1997; 97US-0063350.  
 PR 28-OCT-1997; 97US-0063354.  
 PR 29-OCT-1997; 97US-0063435.  
 PR 29-OCT-1997; 97US-0063704.  
 PR 29-OCT-1997; 97US-0063732.  
 PR 29-OCT-1997; 97US-0063738.  
 PR 29-OCT-1997; 97US-0063734.  
 PR 29-OCT-1997; 97US-0064215.  
 PR 29-OCT-1997; 97US-0063735.  
 PR 31-OCT-1997; 97US-0063870.  
 PR 31-OCT-1997; 97US-0064103.  
 PR 03-NOV-1997; 97US-0064248.  
 PR 07-NOV-1997; 97US-0064809.  
 PR 12-NOV-1997; 97US-0065186.  
 PR 17-NOV-1997; 97US-0065846.  
 PR 18-NOV-1997; 97US-0065693.  
 PR 21-NOV-1997; 97US-0066120.  
 PR 21-NOV-1997; 97US-0066364.  
 PR 24-NOV-1997; 97US-0066772.  
 PR 24-NOV-1997; 97US-0066466.  
 PR 24-NOV-1997; 97US-0066770.  
 PR 24-NOV-1997; 97US-0066511.  
 PR 24-NOV-1997; 97US-0066453.

(GETH ) GENENTECH INC.

PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;

DR WPI; 1999-229533/19.

XX N-PSDB; AAX52213.

PT New isolated human genes and polypeptides used in, e.g. treatment of  
 XX gastrointestinal ulceration

PS Claim 12; Fig 2; 320pp: English.

XX AAY13344-403 represent secreted and transmembrane human proteins.  
 CC The cDNA sequences are obtained from cDNA libraries, prepared from  
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.  
 CC The encoded polypeptides have specific uses based on their homology to  
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders  
 CC associated with the preservation and maintenance of gastrointestinal  
 CC mucosa and the repair of acute and chronic mucosal lesions  
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal  
 CC ulceration and congenital microvillus atrophy), skin diseases associated  
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial  
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),  
 CC potent effects on cell growth and development, diseases related to  
 CC growth or survival of nerve cells including Parkinson's disease,  
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as  
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used  
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment  
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an  
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have  
 CC therapeutic applications in wound healing and tissue repair; PRO317 can  
 CC be used for treating problems of the kidney, uterus, endometrium, blood  
 CC vessels, or related tissue, e.g. in the heart of genital tract.

XX Sequence 353 AA;

Query Match 100.0%; Score 2005; DB 20; Length 353;

Best Local Similarity 100.0%; Pred. No. 1.8e-133;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRLEPRRALGLPLLLPAPPAEAKKPTCHRCGLVDKFNQGVDPATAKNGSGGTAW 60  
 Db 1 mlrlepraalglpllllpapeaakpchrgrglvdkfngmvdctaknfgngltaw 60  
 QY 61 EEKLSKYESESEIRLLELGLCESSDFECNOMLEAOEHLAWMLLKSFPPLFPMFC 120  
 Db 61 eeklskysesseirlllelglelceessdfecnmleaeqehleawmlkseyppllfemfc 120  
 QY 121 VKTLKVCSPGYGPDPDLACQGSQRPCSGNGHSGSGSRQDSCRHMGYQGPLCTDC 180  
 Db 121 vktlkvcspgygpdpdlacqgsqrpcsgnghsgsgsrqdsrchrhmgypplctdc 180  
 QY 181 MDGYFSSLRNETHSICTACDESKCTGSLTRDCGECBVGWLDGACVDVDECAEAPP 240  
 Db 181 mdgyfsslrnethsictadescktcsgltnrdcgeevgwldgacvddvdecaaeapp 240  
 QY 241 CSAAOPCKNANGSYTCBECDSQVCGTGEFGNCKECISGARBRGGCADVDECSLAEXT 300  
 Db 241 csaaopcknangsytcsecdsqvcgctgefgnckecisgarbrggcadvdecslext 300  
 QY 301 CVRKNECYMTPGSGYVCPCDFEETEDACYPAAEATEGESPTQPSREDL 353  
 Db 301 cvrkencytmtpgsgyvcpcdfefeteadacypaaeatgesptqlpsredl 353  
 RESULT 3  
 AAY05283  
 XX AAY05283 standard; Protein; 353 AA.  
 AC AAY05283;  
 XX 22-JUN-1999 (first entry)  
 DE EGF-like homologue PRO211.  
 XX  
 KW Antibody; PRO187; PRO333; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246  
 KW EBAF-2; Inhibitor; Tumour growth; cancer; EGF-like homologue;  
 KW FGF-8 homologue.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9914327-A2.  
 XX 25-MAR-1999.  
 PD  
 XX 10-SEP-1998; 98WO-US18824.  
 XX  
 PR 25-NOV-1997; 97US-0066840.  
 PR 17-SEP-1997; 97US-0059114.  
 PR 17-SEP-1997; 97US-0059117.  
 PR 18-SEP-1997; 97US-0059263.  
 PR 15-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 17-OCT-1997; 97US-0062287.  
 PR 24-OCT-1997; 97US-0062816.  
 PR 29-OCT-1997; 97US-0063704.  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Botstein D, Goddard A, Gurney A, Hillan K, Lawrence DA;  
 PI Roy M, Wood WI;  
 XX WPI; 1999-229533/19.  
 DR N-PSDB; AAX28433.  
 XX  
 PT Antibodies against specific proteins overexpressed in tumours  
 XX  
 PS Example 1; Fig 17; 130pp: English.  
 XX This sequence represents the EGF-like homologue PRO211.



PT Leukemia and for identifying compounds capable of inhibiting  
 PT growth of neoplastic cells  
 XX  
 PS Claim 31: Fig 2, 133pp: English.  
 XX  
 CC Isolated PRO211, PRO228, PRO538, PRO172 or PRO182 polypeptides  
 CC or their agonists (preferably anti-PRO agonist antibody or a small  
 CC molecule mimicking the biological activity of PRO polypeptide) are  
 CC useful in vitro or in vivo for inhibiting the growth of a tumour cell.  
 CC Compositions comprising the PRO polypeptides are useful for  
 CC inhibiting neoplastic cell growth and for treating cancer including  
 CC breast, ovarian, renal, colorectal, uterine, prostate, lung, bladder,  
 CC central nervous system cancer, melanoma and leukaemia in a mammal.  
 CC The PRO polypeptides are also useful for treating other disorders  
 CC such as neuronal, glial, astrocytal, hypothalamic and other glandular,  
 CC macrophagal, epithelial, stromal, blastocellic disorders and  
 CC inflammatory, angiogenic and immunologic disorders as well as being  
 CC useful for identifying agonists to PRO polypeptides by contacting the  
 CC polypeptide with a candidate molecule and monitoring biological  
 CC activity mediated by the polypeptide.  
 CC  
 XX  
 SO Sequence 353 AA:  
 Query Match 100.0%; Score 2005; DB 21; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRLEPRALGLPLLLPLPPAPEAKKPTPCRCGLVDKFNQGMVDTAKKNEGGNTAW 60  
 DB 1 MRLEPRALGLPLLLPLPPAPEAKKPTPCRCGLVDKFNQGMVDTAKKNEGGNTAW 60  
 QY 61 EEKLTSESESEIRLLELGLGESSDFECNOMLEAEHEHLEAWLQLKSPYPLFMFC 120  
 DB 61 EEKLTSESESEIRLLELGLGESSDFECNOMLEAEHEHLEAWLQLKSEYPLFFEW 120  
 QY 121 VKTLKVCSSBGTGPPDCIACOGSGSRPCSGNGHSCGDSRQDSCRCRHMVGQPLCTDC 180  
 DB 121 VKTLKVCSSBGTGPPDCIACOGSGSRPCSGNGHSCGDSRQDSCRCRHMVGQPLCTDC 180  
 QY 121 VKTLKVCSSBGTGPPDCIACOGSGSRPCSGNGHSCGDSRQDSCRCRHMVGQPLCTDC 180  
 DB 121 VKTLKVCSSBGTGPPDCIACOGSGSRPCSGNGHSCGDSRQDSCRCRHMVGQPLCTDC 180  
 QY 181 MDGFFSLRMTHTICTACDESCCTGSLTNRDCECEVGVLDGACVYDDECAAEPPP 240  
 DB 181 MDGFFSLRMTHTICTACDESCCTGSLTNRDCECEVGVLDGACVYDDECAAEPPP 240  
 QY 241 CSAAQFCNNNGSYTCBECSSCGCTGEGPNCKECISGARHREGCAVDDECSLAEXT 300  
 DB 241 CSAAQFCNNNGSYTCBECSSCGCTGEGPNCKECISGARHREGCAVDDECSLAEXT 300  
 QY 301 CVRKENCYNTPGSYVCVCPDGFETEDACVPAEAATEGESPQLPSREDL 353  
 DB 301 CVRKENCYNTPGSYVCVCPDGFETEDACVPAEAATEGESPQLPSREDL 353  
 RESULT 5  
 ID AAY83224 standard; Protein: 353 AA.  
 XX  
 AC AAY83224;  
 XX  
 DT 16-AUG-2000 (first entry)  
 XX  
 DE PRO211 Polypeptide.  
 XX  
 KW Inhibition: cancer; neoplasia; tumour; breast; ovary; renal;  
 KW colorectal; uterus; prostate; lung; bladder; central nervous system;  
 KW CNS; melanoma; leukaemia; PRO211; PRO228; PRO538; PRO172; PRO182;  
 KW human; ss  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..24  
 FT /label= signal\_peptide

FT Modified-site 26..30  
 FT /note= "cAMP and cGMP dependent protein kinase  
 FT phosphorylation site"  
 FT  
 FT Modified-site 44..50  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 54..60  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 55..61  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 58..62  
 FT /note= "Casein kinase II phosphorylation site"  
 FT  
 FT Modified-site 66..70  
 FT /note= "Casein kinase phosphorylation site"  
 FT  
 FT Modified-site 81..87  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 86..90  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 94..116  
 FT /label= "Casein kinase phosphorylation site"  
 FT  
 FT Modified-site 150..156  
 FT /label= "Leucine zipper pattern"  
 FT  
 FT Modified-site 149..153  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 155..159  
 FT /note= "Glycosaminoglycan attachment site"  
 FT  
 FT Modified-site 158..164  
 FT /note= "Glycosaminoglycan attachment site"  
 FT  
 FT Modified-site 164..170  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 166..178  
 FT /note= "N-myristoylation site"  
 FT  
 FT Domain 190..194  
 FT /label= EGF-like domain  
 FT  
 FT Modified-site 197..201  
 FT /note= "N-glycosylation site"  
 FT  
 FT Modified-site 210..214  
 FT /note= "Casein kinase II phosphorylation site"  
 FT  
 FT Modified-site 251..255  
 FT /note= "Casein kinase II phosphorylation site"  
 FT  
 FT Modified-site 252..258  
 FT /note= "N-glycosylation site"  
 FT  
 FT Modified-site 255..259  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 295..299  
 FT /note= "Casein kinase II phosphorylation site"  
 FT  
 FT Modified-site 308..320  
 FT /label= Casein kinase II phosphorylation site  
 FT  
 FT Modified-site 313..319  
 FT /note= "Asn and Asp hydroxylation site"  
 FT  
 FT Modified-site 339..343  
 FT /note= "N-myristoylation site"  
 FT  
 FT Modified-site 349..353  
 FT /note= "Casein kinase II phosphorylation site"  
 FT  
 FT  
 XX  
 XX WO200021996-A2.  
 XX  
 XX PD 20-APR-2000.  
 XX  
 XX XX 05-OCT-1999; 99WO-US23089.  
 XX  
 XX PR 13-OCT-1998; 98US-0104080.  
 XX  
 XX  
 PA (GENTECH ) GENENTECH INC.  
 XX  
 XX Ashkenazi A, Goddard A, Gurney AL, Klein RD, Napier M, Wood WI,  
 PI Yuan J;  
 XX  
 DR WPI: 2000-317943/27.  
 DR N-PSDB: AA293700.  
 XX  
 XX Composition for inhibiting neoplastic cell growth and treating cancers  
 PT of ovary, uterus, prostate, lung and bladder, comprises PRO211, PRO228,  
 PT PRO538, PRO172 or PRO182 polypeptide or their agonist

XX Claim 5; Figure 2; 122pp; English.  
 PS Compositions comprising a PRO211, PRO228, PRO538, PRO172 or PRO182  
 CC polypeptide or their agonists, mixed with a carrier is useful for  
 CC inhibiting neoplastic growth and treating tumors such as cancers of  
 CC breast, ovary, renal, colorectal, uterus, prostate, lung, bladder,  
 CC central nervous system, melanoma and leukaemia.  
 XX  
 SQ Sequence 353 AA:

Query Match 100.0%; Score 2005; DB 21; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRRAALGLPLLLLPAPPAKAPPCPHRCRLVDFKNGWMDTAKKNGGGNTAW 60  
 DB 1 mrlpraaalglpllllpapeaakpchrclrvdfkngwmdtakknfggntaw 60  
 QY 61 EKKTSKYESSERILEILEGLCESSDFECNQMLEAQEHLFAWMLQLKSEYDLEFEMFC 120  
 DB 61 eekltskysesserilleileglcessdfecnqmlaeqehlfaawmlqlkseypdlfetc 120  
 QY 121 VRTLVKVCSPGYTPDCLACGGGSGRPGCSGNHCSGDSGRQDGSRCRMGYQGPICTDC 180  
 DB 121 vrtlvkcspgytpdclacggsgprpcsgnhcsgdsgsrqdgsrcrmgyqgpictdc 180  
 QY 181 MGVFSLNENHSTCTACDESKTCSGLTNRDCGEVGMVLTDEGACVDVDECAEPP 240  
 DB 181 mgyfslnenhstctacdesktcsgltnrdcgevwmvltdegacvdrvdecaep 240  
 QY 241 CSAAPFCNKANSGYTCEDSCVCGTGEAPGNCKECISGYAREHQACADVDECSIAEKT 300  
 DB 241 csaapfcnkansgytcedscvctgeapgnckecisgyarehqacadvdecsiaekt 300  
 QY 301 CVRKNECYNTPGSYVCPDGEFETEDACVPAPAEATGESPTQLPSREDL 353  
 DB 301 cvrkencyntpgsyvcpdgefetedacvpapaeatgesptqlpsredl 353

RESULT 6  
 AAU12319  
 ID AAU12319 standard; Protein; 353 AA.  
 AC AAU12319;  
 XX  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Human PRO211 polypeptide sequence.  
 XX  
 KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;  
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;  
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;  
 KW adipocyte; A-peptide; factor VIIA; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200140466-A2.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 01-DEC-2000; 2000WO-US32678.  
 XX  
 PR 01-DEC-1999; 99WO-US28301.  
 PR 01-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28564.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 09-DEC-1999; 99US-0170262.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US30911.  
 PR 20-DEC-1999; 99WO-US30999.

PR 30-DEC-1999; 99WO-US31243.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 PR 11-FEB-2000; 2000WO-US03565.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 22-FEB-2000; 2000WO-US04414.  
 PR 24-FEB-2000; 2000WO-US05004.  
 PR 01-MAR-2000; 2000WO-US05601.  
 PR 20-MAR-2000; 2000WO-US07377.  
 PR 21-MAR-2000; 2000WO-US07532.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 10-NOV-2000; 2000WO-US30873.  
 XX  
 PA (GEN) GENENTECH INC.  
 PI Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;  
 PI Gerlitsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
 PI Smith V, Stewart RA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
 XX WPI: 2001-408281/43.  
 DR N-PSDB; AAS21391.  
 XX  
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect  
 PT other PRO polypeptides, link bioactive molecules to cells expressing  
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.  
 PT lung, breast, prostate, cervical  
 XX  
 PS Claim 12; Fig 296; 813pp; English.  
 XX  
 CC AAU12172-AAU12446 represent novel human secretory and transmembrane  
 CC PRO polypeptides. The PRO polypeptides are useful to detect other  
 CC PRO polypeptides, to link bioactive molecules to cells expressing  
 CC PRO polypeptides, to modulate biological activities of cells expressing  
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,  
 CC breast, prostate, cervical or liver tumours by comparing PRO  
 CC polypeptide expression in a cell sample to that in a control sample.  
 CC Some of the 275 sequences are also useful to stimulate the release of  
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the  
 CC proliferation or differentiation of chondrocytes, the proliferation or  
 CC gene expression in pericyte cells, the release of proteoglycans from  
 CC cartilage, the proliferation of inner ear utricular supporting cells or  
 CC of T-lymphocytes, the release of a cytokine from peripheral blood  
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of  
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by  
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide  
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify  
 CC molecules involved in binding interactions. The polynucleotides encoding  
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,  
 CC transgenic or knock out animals and can be used in gene therapy.  
 XX  
 SQ Sequence 353 AA:

Query Match 100.0%; Score 2005; DB 22; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRRAALGLPLLLLPAPPAKAPPCPHRCRLVDFKNGWMDTAKKNGGGNTAW 60  
 DB 1 mrlpraaalglpllllpapeaakpchrclrvdfkngwmdtakknfggntaw 60  
 QY 61 EKKTSKYESSERILEILEGLCESSDFECNQMLEAQEHLFAWMLQLKSEYDLEFEMFC 120  
 DB 61 eekltskysesserilleileglcessdfecnqmlaeqehlfaawmlqlkseypdlfetc 120  
 QY 121 VRTLVKVCSPGYTPDCLACGGGSGRPGCSGNHCSGDSGRQDGSRCRMGYQGPICTDC 180  
 DB 121 vrtlvkcspgytpdclacggsgprpcsgnhcsgdsgsrqdgsrcrmgyqgpictdc 180

Db 121 vktlkvcspgtygpdclacqgsqrgpsngbhsqdsrgdscrchmgyp99plctdc 180  
 QY 181 MDGFSSLRNETHSICACDESKCTGSLTRDCECEVGVNLDGACVVDDECAAEPP 240  
 Db 181 mdgylsslrnehtsictadescktsqllnrddgecevgvnldegacvvddecaaepp 240  
 QY 241 CSAAOFCKNANGSYTCESDSSCVGCTGEGPGNCKECISGYAREHGOCADVDECSLAET 300  
 Db 241 csaaqfcknangsytccecdsscvgctggegpgnckecisgyarehgocadvdecslaekt 300  
 QY 301 CVRKNENCYNTPGSYVCVCPDGFETEDACVPAAEAETEGESPTQLPSREDL 353  
 Db 301 cvrknenctnpgsyvcvcpdgfeetedacvpaaeaetegesptqlpsredl 353

## RESULT 7

AAB68596 standard; Protein: 353 AA.

AC AAB68596;

DT 27-APR-2001 (first entry)

DE PRO211.

KM Cytostatic; PRO protein; tumour; cancer.

OS Homo sapiens.

PN WO200105836-A1.

PD 25-JAN-2001.

PF 20-DEC-1999; 99WO-US30999.

PR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.

PR 08-SEP-1999; 99WO-US20594.

PR 13-SEP-1999; 99WO-US20944.

PR 15-SEP-1999; 99WO-US21090.

PR 05-OCT-1999; 99WO-US23089.

PR 29-NOV-1999; 99WO-US28214.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28564.

PA (GETH ) GENENTECH INC.

PI Botstein D, Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI;

DR WPI: 2001-091968/10.

DR N-PSDB: AAF60360.

PT New antibody that binds to a PRO polypeptide, e.g. PRO187 and PRO533,

PS useful for diagnosing and treating cancers -

PS Claim 61: Fig 10; 196pp; English.

CC The present invention relates to PRO proteins and coding sequences. The

CC present sequence is one such PRO protein. It was found that the PRO genes

CC are amplified in the genome of tumour cells. The gene amplification is

CC expected to be associated with the overexpression of the gene product and

CC contributes to tumorigenesis. Therefore, antagonists of PRO proteins are

CC useful for the treatment of benign or malignant tumours, leukaemias,

CC lymphoid malignancies and other disorders such as neuronal, glial,

CC astrocytic, hypothalamic, glandular, epithelial, inflammatory and

CC immunologic disorders.

XX Sequence 353 AA;

QY 1 MRLPRRALGLPLPLLLPPAPEAKKPTPCRRCGLVDKPNOCMDVPTAKNFGGNTAW 60  
 Db 1 mrlprraiglpllllppapeaakkptpcrcrglvdkngmvdtaaknfggntaw 60  
 QY 61 EEKTLKSYESSEIRLLELEGLCESDSEFCNOMLEAOEHLAAMWLKSEYPLDFMFC 120  
 Db 61 eektlksyesselrllleleglcesdsefcnmleaqehleawllkseypldfmfc 120  
 QY 121 VKTLLKVCSPGTYGPDCLACGGSQRPCSGNGHSGGSGRGGSCRCRHMCTGCPCLTDC 180  
 Db 121 vktlkvcspgtygpdclacqgsqrgpsngbhsqdsrgdscrchmgyp99plctdc 180  
 QY 181 MDGFSSLRNETHSICACDESKCTGSLTRDCECEVGVNLDGACVVDDECAAEPP 240  
 Db 181 mdgylsslrnehtsictadescktsqllnrddgecevgvnldegacvvddecaaepp 240  
 QY 241 CSAAOFCKNANGSYTCESDSSCVGCTGEGPGNCKECISGYAREHGOCADVDECSLAET 300  
 Db 241 csaaqfcknangsytccecdsscvgctggegpgnckecisgyarehgocadvdecslaekt 300  
 QY 301 CVRKNENCYNTPGSYVCVCPDGFETEDACVPAAEAETEGESPTQLPSREDL 353  
 Db 301 cvrknenctnpgsyvcvcpdgfeetedacvpaaeaetegesptqlpsredl 353

## RESULT 8

AAB80212 standard; Protein: 353 AA.

AC AAB80212;

DT 24-APR-2001 (first entry)

DE Human PRO211 protein.

XX Human: PRO; dermatological; antipruritic; cytostatic; antiinflammatory;

XX antiparkinsonian nootropic; neuroprotective; vulnerrary; candid;

XX antiangiogenic; vasotropic; antiaesthetic; antineumatic; cancer;

XX antiarthritic; antidiabetic; antilivral; diabetes;

XX ophtalmological; gene therapy; skin disease; gastrointestinal disorder;

XX ischaemia; inflammation.

OS Homo sapiens.

PN WO200104311-A1.

PD 18-JAN-2001.

PF 22-FEB-2000; 2000WO-US04414.

PR 07-JUL-1999; 99US-0143048.

PR 26-JUL-1999; 99US-0145698.

PR 28-JUL-1999; 99US-0146222.

PR 08-SEP-1999; 99WO-US20594.

PR 13-SEP-1999; 99WO-US20944.

PR 15-SEP-1999; 99WO-US21090.

PR 05-OCT-1999; 99WO-US21547.

PR 29-NOV-1999; 99WO-US23089.

PR 30-NOV-1999; 99WO-US28214.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 05-JAN-2000; 99WO-US00219.

PA (GETH ) GENENTECH INC.

PI Botstein D, Desnoyers L, Eaton DL, Ferrara N,

PI Filvaroff E, Fong S, Gao W, Garber H, Gerritsen ME, Goddard A;

PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavlin I;

PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;

PI Williams PM, Wood WI;

XX WPI; 2001-081051/09.  
 DR N-PSDB; AAF72371.  
 XX  
 PT Sixty one nucleic acids encoding PRO polypeptides which are useful in  
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung  
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.  
 PT Alzheimer's disease) -  
 XX  
 PS Claim 1; Fig 2; 393pp; English.  
 XX  
 CC The present sequence is one of sixty one novel secreted and  
 CC transmembrane PRO polypeptides. The PRO polypeptides are  
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung  
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.  
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.  
 CC endometrial bleeding, angiogenesis, ischaemia such as coronary  
 CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,  
 CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and  
 CC diabetes and retinal disorders such as retinitis pigmentosa.  
 CC The PRO nucleic acids have applications in molecular biology, including  
 CC use as hybridization probes, and in chromosome and gene mapping.  
 CC  
 SQ Sequence 353 AA:

Query Match 100.0%; Score 2005; DB 22; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRRAALGLPLLLLPAPPAEAKKPPCHRCRGLVDFKNGWVDTAKKNGGNTAM 60  
 DB 1 mlprtraalglpllllpapeaakkpchrhcrglvdfkngwvdtakkngfgntaw 60  
 QY 61 EKKTLISKYSSERLLEILLEGICSSDFECNOMLEAEOHELEAWMLQKSEYFDLFEWFC 120  
 DB 61 eektliskysseirillellegicessdfecnmlaeeheleawmlqlkseyfdlfewfc 120  
 QY 121 VKTLKWCSPGTYGPDCCLACOGGSGORPCSGNGHCSGDSGROGSGRCRHHGVPGLCTDC 180  
 DB 121 vktlkwcspgtygpdcclacoggsgorpcsgngchscgdsgrgdsgrchmgygplctdc 180  
 QY 181 MDGYFSLRNETHSICTACDESCKTCGSLTNRDGCCEVGMWLDGACVYDVECAAEPP 240  
 DB 181 mdgyfslrnethsictacdescctcsgltnrdgccevgmwldgacvdydecaaepp 240  
 QY 241 CSAAOCCKNANGSYTCEEDSSCVGCTGEGPNCRCICIGYAREHOCADYDVECSLAEXT 300  
 DB 241 csaagfcknangsyteeedsscvgctgegpncrcicisgyarehcgadvedecsiaext 300  
 QY 301 CVRKNENCYNTPGSYVCVPDGFEEFEDACVPAEAATEGESPTQLPSREDL 353  
 DB 301 cvrkencyntpgsyvcvpdgffeedacvpaaeaategesptqlpsredl 353

RESULT 9  
 AAB61231  
 ID AAB61231 standard; Protein; 353 AA.  
 XX  
 AC AAB61231;  
 XX  
 DT 03-APR-2001 (first entry)  
 XX  
 DE Human TANGO 331 protein.  
 XX  
 KW Human; INTERCEPT 217; INTERCEPT 297; TANGO 276; TANGO 292; TANGO 325;  
 KW TANGO 331; TANGO 332; cytosolic; antiinflammatory; antiarrhythmic;  
 KW antiapoptotic; gene therapy; cancer; inflammatory disorder;  
 KW cardiac disorder; arrhythmia; skin disorder; psoriasis.  
 XX  
 OS Homo sapiens.  
 XX

PN 16-JUN-2000; 2000WO-US16658.  
 XX  
 PD 16-JUN-2001.  
 XX  
 PR 29-JUN-1999; 99US-0342364.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Kirst SJ, Holtzman DA, Fraser CC, Sharp JD, Barnes TS;  
 XX  
 DR WPI; 2001-061966/07.  
 DR N-PSDB; AAF29457, AAF29458.  
 XX  
 XX

Isolated human proteins are used for diagnosis, treatment and prevention of cancers, inflammatory disorders, cardiac disorders e.g. arrhythmia, and skin disorders e.g. psoriasis -

Claim 8; Fig 6A-6C; 372pp; English.

The present sequence is given in a specification relating to isolated human proteins designated INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331 and TANGO 332. These proteins are useful as modulating agents or as targeting agents for developing agents to regulate cellular processes e.g. growth, proliferation, survival, differentiation and activity of human tissues. Diseases which can be diagnosed, prevented and treated by administration of these polypeptides; their nucleic acids and modulators include cancers, inflammatory disorders, cardiac disorders e.g. arrhythmia, and skin disorders e.g. psoriasis. Nucleic acids encoding the isolated proteins can be used to express the proteins in a host cell in gene therapy applications. Antisense molecules or ribozymes can be used to inhibit expression of the proteins in target cells. Fragments of the nucleic acid molecules encoding the isolated proteins are used as hybridisation probes and as polymerase chain reaction (PCR) primers.

Query Match 100.0%; Score 2005; DB 22; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRRAALGLPLLLLPAPPAEAKKPPCHRCRGLVDFKNGWVDTAKKNGGNTAM 60  
 DB 1 mlprtraalglpllllpapeaakkpchrhcrglvdfkngwvdtakkngfgntaw 60  
 QY 61 EKKTLISKYSSERLLEILLEGICSSDFECNOMLEAEOHELEAWMLQKSEYFDLFEWFC 120  
 DB 61 eektliskysseirillellegicessdfecnmlaeeheleawmlqlkseyfdlfewfc 120  
 QY 121 VKTLKWCSPGTYGPDCCLACOGGSGORPCSGNGHCSGDSGROGSGRCRHHGVPGLCTDC 180  
 DB 121 vktlkwcspgtygpdcclacoggsgorpcsgngchscgdsgrgdsgrchmgygplctdc 180  
 QY 181 MDGYFSLRNETHSICTACDESCKTCGSLTNRDGCCEVGMWLDGACVYDVECAAEPP 240  
 DB 181 mdgyfslrnethsictacdescctcsgltnrdgccevgmwldgacvdydecaaepp 240  
 QY 241 CSAAOCCKNANGSYTCEEDSSCVGCTGEGPNCRCICIGYAREHOCADYDVECSLAEXT 300  
 DB 241 csaagfcknangsyteeedsscvgctgegpncrcicisgyarehcgadvedecsiaext 300  
 QY 301 CVRKNENCYNTPGSYVCVPDGFEEFEDACVPAEAATEGESPTQLPSREDL 353  
 DB 301 cvrkencyntpgsyvcvpdgffeedacvpaaeaategesptqlpsredl 353

RESULT 10  
 AAB53075  
 ID AAB53075 standard; Protein; 353 AA.  
 XX



AC AAB53075;  
 XX 28-FEB-2001 (first entry)  
 XX  
 DE Human angiogenesis-associated protein PRO211, SEQ ID NO:57.  
 XX  
 KW Human; angiogenesis-associated protein; PRO; endothelial cell growth;  
 KW cardiac hypertrophy; cardiovascular disorder; endothelial disorder;  
 KW angiogenic disorder; atherosclerosis; osteoporosis; hypertension;  
 KW amocardial infarction; diabetic retinopathy; rheumatoid arthritis;  
 KW Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;  
 KW Alzheimer's disease; Huntington's disease; stroke; drug screening;  
 KW gene therapy; transgenic animal.  
 XX  
 OS Homo sapiens.  
 XX  
 PN NO200053753-A2.  
 PD 14-SEP-2000.  
 XX  
 PF 05-JAN-2000; 2000MO-US00219.  
 XX  
 PR 08-MAR-1999; 99MO-US05028.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 02-JUN-1999; 99MO-US12252.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 01-SEP-1999; 99MO-US20111.  
 PR 08-SEP-1999; 99MO-US20594.  
 PR 15-SEP-1999; 99MO-US21090.  
 PR 15-SEP-1999; 99MO-US21547.  
 PR 05-OCT-1999; 99MO-US23089.  
 PR 30-NOV-1999; 99MO-US28313.  
 PR 30-NOV-1999; 99MO-US28409.  
 PR 02-DEC-1999; 99MO-US28564.  
 PR 02-DEC-1999; 99MO-US28565.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Goddard A;  
 PI Godowski PJ, Gurney AL, Hillan KJ, Kuo SS, Mark MR, Marsters SA;  
 PI Pooni NF, Pittl RM, Watanabe CK, Williams PM, Wood WI;  
 XX  
 XX WPI: 2001-090793/10.  
 DR N-PSDB; AAC97409.  
 XX  
 XX New isolated nucleic acid for producing a PRO polypeptide, analyzing  
 XX genetic disorders and treating cardiovascular, endothelial or  
 PT angiogenic disorders, such as atherosclerosis, wounds or cancer -  
 PT  
 XX Claim 69, Fig 24; 293pp; English.  
 PS  
 XX The invention relates to novel human angiogenesis-associated proteins  
 CC designated PRO proteins (AAB53064-B53097), and to nucleic acids encoding  
 CC PRO proteins. The invention also relates to vectors and host cells  
 CC comprising a PRO nucleic acid, the recombinant production of a PRO  
 CC protein, PRO antibodies specific for a PRO protein, fusion proteins  
 CC comprising a PRO protein, agonists or antagonists of a PRO protein, and  
 CC compounds which inhibit the expression of a PRO gene. The invention  
 CC additionally encompasses methods of identifying modulators of PRO  
 CC expression or activity; diagnosing a cardiovascular, endothelial or  
 CC angiogenic disorder, or a susceptibility to such a disorder by detecting  
 CC mutations in a PRO gene, or the expression level of a PRO gene within a  
 CC particular tissue; treating a cardiovascular, endothelial or angiogenic  
 CC disorder via the administration of a PRO protein, PRO nucleic acid, or  
 CC PRO agonist or antagonist; a retroviral gene therapy vector comprising a  
 CC PRO nucleic acid; and methods of inhibiting or stimulating endothelial  
 CC cell growth, cardiac hypertrophy or PRO-induced angiogenesis via the  
 CC administration of a PRO protein, or an agonist or antagonist thereof.  
 CC PRO nucleic acids, PRO proteins, antibodies against PRO proteins, PRO  
 CC agonists and PRO antagonists may be used as therapeutic agents to treat

CC cardiovascular, endothelial or angiogenic disorders, such as  
 CC atherosclerosis, osteoporosis, myocardial infarction, hypertension,  
 CC diabetic retinopathy, rheumatoid arthritis, Crohn's disease, psoriasis,  
 CC endometriosis, ulcers, wounds, cancer, Alzheimer's disease, Huntington's  
 CC disease, or stroke. PRO nucleic acids are additionally useful in the  
 CC recombinant production of PRO proteins, as hybridisation probes to  
 CC screen libraries to isolate cDNAs with sequence identity to PRO proteins;  
 CC to map genes encoding PRO proteins, to analyse genetic disorders, and in  
 CC gene therapy. PRO nucleic acids can also be used to produce transgenic  
 CC animals useful for the development and screening of potential  
 CC therapeutic agents. The present sequence represents a PRO protein of the  
 CC invention.  
 CC  
 XX  
 XX Sequence 353 AA;  
 XX  
 SO  
 Query Match 100.0%; Score 2005; DB 22; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1,8e-133;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRLPRRAALGILPLLLPPAPEAKKPTPOHRCGLVDRKPNQGMVDTAKKNGCGNTAM 60  
 Db 1 mrlprraalglpllllppapeaakptpohrcrglvdrkpnqgmvdtaakniggnltaw 60  
 QY 61 EEKTLSEYSESEIRLEILLEGICSDPEECNQMLEAOEHLAAMWLKSEYPLPFMFC 120  
 Db 61 eektlseyseseirleilelegicsdpeecnmleaoehleawwlkseyplpfefwc 120  
 QY 121 VKTLKVCSPGYTPDCLACOGSGSRPCSGNGHSGSGSRGDSGSRCHMGYGPICITDC 180  
 Db 121 vktlkvcspgytpdc lacqgsqrpcsgngshsgsrqgdsgrchmgygplctdc 180  
 QY 181 MDGYFSSLRNETHSICRNACDESKTCGTLTRDGCBEVGVNLDGACVYDDECAEAPP 240  
 Db 181 mdgyfsslrnethsictacdescktcsgtltrdgcbevgnldgacvddcaaeapp 240  
 QY 241 CSAAPFCNANGSYTCECDSYCVCTGEGPGNCKECISGARHRCACADVDECSLAEXT 300  
 Db 241 csaaqfcnkangsytccecdssvcvctgepgnckecisgarhrgcacadvdecslekt 300  
 QY 301 CVRKNEHCYNTPGSYVCVCPDGFETEDACVPAEAETEGESPTQLPSREDL 353  
 Db 301 cvrknehcynypgsyvcvcpdgfeetedacvpaeaetegesptqlpsredl 353  
 RESULT 11  
 AAY8571  
 ID AAY8571 standard; Protein: 353 AA.  
 XX  
 XX AAY8571;  
 AC  
 XX 09-AUG-2000 (first entry)  
 DT  
 XX  
 XX Human PRO211 amino acid sequence.  
 DE  
 XX Antibody: PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246  
 KW PRO317; tumour growth inhibitor; cancer; diagnosis; treatment; human;  
 KW cell growth; proliferation; epidermal growth factor; EGF; ADGFP;  
 KW antibody dependent enzyme mediated prodrug therapy.  
 OS  
 XX Homo sapiens.  
 OS  
 PN WO200015666-A2.  
 PD 23-MAR-2000.  
 PD  
 XX 08-SEP-1999; 99MO-US20594.  
 PF  
 XX 10-SEP-1998; 98US-0099803.  
 PR 10-SEP-1998; 98MO-US18824.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA

Pt	GoddardA.	Gurney AL,	Hillan KJ,	Roy MA,	Wood WI,	Botstein D;
xx						
DR	WPt:	2000-271386/23.				
xx	N-PSDB:	AAA30040.				
Pt	New isolated antibodies which bind to specific polypeptides used for					
Pt	diagnosis and treatment of neoplastic cell growth and proliferation -					
xx						
xx	Example 5; Fig 10; 200pp; English.					
CC	This sequence represents a human PRO211 amino acid sequence. PRO211					
CC	shares sequence homology with the epidermal growth factor protein					
CC	sequence. The invention relates to isolated antibodies which bind to a					
CC	polypeptide. The "PRO" polypeptides are encoded by genes which are over					
CC	expressed in the genome of tumour cells. Vectors and host cells					
CC	comprising the nucleic acid encoding the antibodies are used in the					
CC	production of the antibodies. The antibodies and nucleic acids encoding					
CC	them are used for diagnosing a tumour in a mammal. The antibodies are					
CC	used for inhibiting the growth of tumour cells and identifying compounds					
CC	that inhibit a biological or immunological activity of and/or expression					
CC	of a PRO187, PRO333, PRO214, PRO240, PRO211, PRO230, PRO261, PRO246 or					
CC	PRO317 polypeptide. The antibody may be used in antibody dependent enzyme					
CC	mediated prodrug therapy (ADPEP) by conjugating the antibody to a					
CC	prodrug-activating enzyme which converts a prodrug to an anti-cancer					
CC	drug. The antibodies can be fluorescently labelled and monitored by light					
CC	microscopy, flow cytometry or fluorimetry for diagnosis and prognosis of					
CC	tumours.					
xx						
SQ	Sequence	353 AA:				
	Query Match	99.6%;	Score 1997;	DB 21;	Length 353;	
	Best Local Similarity	99.7%;	Pred. No. 6.5e-133;			
	Matches 352;	Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0
OY	1 MFLPPRAALGLPLILLILPPAPEAAKKPRPHRCRGELVKNCGMWDATKKNFGGNTAM	60				
Db	1 mlplpraaagllpillailllppepaakkpchpcnrcgildvknqgmvtlakkmgggntaw	60				
OY	61 EERTLSKYESSERIRLEIIEGLCESSDFECNOMLEAOEHLLEAWMLQLSEYPDLEWMC	120				
Db	61 eektlskyessetrillellejglccesdfecnmqlaeqeelaeawlqlkseydilewfc	120				
OY	121 VTLTWCCSPGRGYGPDCACOGGSQRPGSGNGHSCSDSGRGGGSCRCHMGYGPICTOC	180				
Db	121 vcltlvwccspgygpdcacoggsqrpgsgngncsgdsrgdgscrchmgygpiictdc	180				
OY	181 MDGVSSSLNETHSICTACDESKCKTCSGTNRDGCCEEWGWVLDDEACADVDECAAEPPP	240				
Db	181 mdgvssslnehtsiactadesckctcsgtlnrdgcceewgwvldegacvdvedcaaeppp	240				
OY	241 CSAAPFCRKANASYICEEDSSCVGTGSGPKGKCICIGYAREHQCADVDICSIAEKT	300				
Db	241 csaafpfckmansyiceedsscvcvtgsgpkkcicisgyarehqcadvdicsiaektt	300				
OY	301 CVRKHNENCNTPGSGSYVCVPDGFEEETEDACVPAAEAATEGESPTOLPSREDL	353				
Db	301 cvrkhnencnptpgsgsyvcvpdgdfeetledacvpaaeaategesptqlpsredl	353				
RESULT 12						
ID	AAB61233					
XX	AAB61233 standard; Protein: 329 AA.					
XX						
AC	AAB61233:					
XX						
XX	03-APR-2001 (first entry)					
XX						
DE	Mature human TANGO 331 protein.					
KW	Human: INTERCEPT 217; INTERCEPT 297; TANGO 276; TANGO 292; TANGO 325;					
KW	TANGO 331; TANGO 332; cytoskeletal; antinflammatory; antiarrhythmic;					
KW	antiproliferic; gene therapy; cancer; inflammatory disorder;					

Query Match	Best Local Similarity	Matches 329: Conservative	94.1%: 100.0%: 0: Mismatches	Score 1887: Pred. NO. 3.4e-125; 0: Indels	DB 22: Length 329: 0: Gaps
QY 25 AKKPRCHRCRLGVKFNQGWMTAKKKNFGGNGNTAMEEKTLSKSYESSSEIRLLEILEGLCE 84					
Db 1 akkprchrcrglvdkfngmwdtaknfigntaweektlakysessellleilegle 60					
QY 85 SSDFEFCNOMLEAEOEHLLEAMVQLKSEFIDPFEWFCVKTLLKCCSPGTYGPPCLACGGGS 144					
Db 61 ssdfecncmlleaqeehleawvlgklsyepdflfewfcvktllkccspgtygppclacgggs 120					
QY 145 ORPCGNGCHSGDGSROGSGSCRCMAGVGGPICTGCMODGYFSFLNRTFSTICTADDESK 204					
Db 121 qrpccngchscdgsrvgdgsrccrmagvggpiictcmodyfsslnrthstictadesck 180					
QY 205 TCSGLTNRDCEGECEGWLDEGACVVDVECAAEPPCSAAQFCCKNANGSYTCECDSSCV 264					
Db 181 tcsgltnrdcegecegwlddegacvvdvecaaeppcsaaqfccknangsyteeccsscv 240					
QY 285 GCTGGGPRNGKRCISIGYAREHGQCADVDVBCSLAEKTCVRKNNKNCVNTGOSYCVCPDPGE 324					
Db 241 gctgggprngkrcisigyarehgqcadvdbcslaektcvrrknnkncvntgositycvcpdpge 300					
QY 325 ETEDACVPPAEAEATGESPTQLPARED 353					
Db 301 etedacvppaeaeategesptqlpared 329					

AAB42711  
 ID AAB42711 standard; Protein; 318 AA.  
 XX  
 AC AAB42711;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF2475 polypeptide sequence SEQ ID NO:4950.  
 XX  
 KW Human: open reading frame: ORFX: detection: cytosolic; hepatotropic;  
 KW vulnerable; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;  
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 KM  
 XX Homo sapiens.  
 OS  
 XX MO200058473-A2;  
 PN  
 XX PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000MO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 DR N-PSDB; AAC76920.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 4122-4123; 5507pp; English.  
 XX  
 CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerable;  
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antirheumatic; antihypertensive;  
 CC antinaemic; The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SO Sequence 318 AA;

Query Match 85.6%; Score 1716; DB 21; Length 318;  
 Best Local Similarity 97.4%; Pred. No. 3.5e-113;  
 Matches 298; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 20 PAPAARPPPCRRCLGVNFKFNQGNVDTAKKNFGGNTAAEKKTLSESEIRLLEIL 79  
 Db 2 paacpnyltapehrcrvlvdkfngmvdtaaknfggntaweektlkysseirllleil 61  
 QY 80 EGLCESSDFECNDMLLEAQEHLLEAMWLQKSEYPDLEFMEVCVTKLKVCSPGTGPPCLA 139  
 Db 62 eglcessefdeconplleaeqeellleawwllkseyddlfevfkllkvcspgtygpclla 121  
 QY 140 CQGSORPCGNGNCSGDSNRQGSRCRHMGVGPRLCTDCMOCYFESSLRNERTSTTAC 199  
 Db 122 cqqsgqrpccngnscsgdsnrqgsrchrhmgvgprlctdcmdgyfisslrnertsttacc 181  
 QY 200 DESCKTSGLTNRDCCGCEVGWVLDGACVVDVDECAAEPPPCSAAPCCKNANGSYTCEEC 259  
 Db 182 descktsqnlndcgecevgwvldgacvvdvdecaaepppcsaagfcknangsytceec 241  
 QY 260 DSSCVGCTGEGPCNCKECISGAYAREHGQCADVDECSLAERTCVKKNENCTPESYVCV 319  
 Db 242 dsscvgtcgegpncnckecisgayarehgqcadvdecslaektcvkrkenctpessyvcv 301  
 QY 320 PDGFE 325  
 Db 302 pdgfee 307

RESULT 14  
 AAY91870  
 ID AAY91870 standard; Protein; 353 AA.  
 XX  
 AC AAY91870;  
 XX  
 DT 19-JUL-2000 (first entry)  
 XX  
 DE Human apoptosis related protein.  
 XX  
 KW Human apoptosis related protein; ARP; modulator; proliferative; cancer;  
 KW programmed cell death; neurodegenerative; Alzheimer's Disease.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 140..330  
 FT /label= EGF\_EGF-1like\_domain  
 FT Domain 148..187  
 FT /label= Laminln\_EGF-1like\_domain  
 FT Misc-difference 190  
 FT /label= OTHER  
 FT /note= "OTHER - not defined in the specification,  
 encoded by GRA"

MO200012236-A2;  
 30-MAR-2000.  
 24-SEP-1999; 99MO-US22270.  
 24-SEP-1998; 98US-0160623.  
 (MILL-) MILLENNIUM PHARM INC.  
 Rhodadoust MM;  
 WPI; 2000-283545/24.  
 DR N-PSDB; AAA08503, AAA08504.  
 New apoptosis related proteins and nucleic acid molecules used for  
 regulating cellular processes e.g. programmed cell death -

XX PS Claim 2; Page 99-100; 102pp; English.

XX CC The present sequence is a novel human apoptosis related protein (hARP).

XX CC The hARP was isolated from a human heart cDNA library. Positive

CC clones were isolated and subsequently 5' RACE PCR was used to obtain the

CC full length clone. ARP and the nucleic acid molecules encoding ARP are

CC used as modulating agents for regulating cell processes, e.g. associated

CC with proliferative disorders such as cancer and neurodegenerative

CC disorders such as Alzheimer's Disease. They can be used

CC in screening assays, predictive medicine and methods of treatment. ARP

CC can be used to modulate interaction with a non-ARP protein molecule on

CC the surface of the same cell which expresses it, modulate interaction

CC with a non-ARP protein molecule on the surface of a different cell,

CC activate an ARP-dependent signal transduction pathway and modulate

CC programmed cell death. The nucleic acids can be used to express ARP

CC protein, detect ARP mRNA or a genetic alteration in an ARP gene and

CC modulate ARP activity. The vector and host cell may be used to generate

CC transgenic animals. Peptide nucleic acids (PNAs) of ARP nucleic acid

CC molecules can be used as antisense or antigenic agents for

CC sequence-specific modulation of gene expression by inducing transcription

CC or translation arrest or inhibiting replication. They can also be used in

CC the analysis of single base pair mutations in a gene or as probes for DNA

CC sequencing or hybridization.

XX SQ Sequence 353 AA;

Query Match 82.1%; Score 1647; DB 21; Length 353;  
 Best Local Similarity 83.6%; Pred. No. 2.8e-108;  
 Matches 306; Conservative 6; Mismatches 28; Indels 26; Gaps 4;

QY 1 MRLPRRAALGLPLLLLPAPAPAKKPPCHRCRGLVNDKFNQGNVDTAKKNFGGNTAM 60  
 1 mlprraalglpllllpapaakppchrcrglvndkfngmntakknfggntaw 60

QY 61 EEXTLSKYESSERLRLIEGLDESSDFECNOMLEQNEHLEAMWLQKSEYFDLEFMC 120  
 61 eextlskyesserlrlliegldessdfecnomleqnehleamwlqkseyfdlfevic 120

QY 121 VKTLKCCSPGTGTPCLOCGGSGRSCNGHSGDSHOGSGSCRHMGYGPPLCTPC 180  
 121 vktlkccspgtgtptclccggsggrscngshsgdshogsgscrhmgypplctpc 180

QY 121 VKTLKCCSPGTGTPCLOCGGSGRSCNGHSGDSHOGSGSCRHMGYGPPLCTPC 180  
 121 vktlkccspgtgtptclccggsggrscngshsgdshogsgscrhmgypplctpc 180

QY 181 MDGFFSLRNETHSICTACDESCRTCSGLTRNDGCEGEVGLDE--GACVDVDEC-- 234  
 181 mdgffslrnethsictacdescrtcsgltrndgcegevwldede--gacvdydec-- 234

QY 181 mdgffslrnethsictacdescrtcsgltrndgcegevwldede--gacvdydec-- 234  
 181 mdgffslrnethsictacdescrtcsgltrndgcegevwldede--gacvdydec-- 234

QY 235 -----AAEPFCSAAGCKNANNGSYTCEECSSCVGCTGEBPGCKECSISYAREHQ 287  
 235 -----aaepfcsaagcknannngsytyceecsscvgctgebpckecsisyarehq 287

QY 228 apywmwtsvrtptvpaacqfcknpngsytyceecsscvgctgebpckecsisyarehq 287  
 228 apywmwtsvrtptvpaacqfcknpngsytyceecsscvgctgebpckecsisyarehq 287

QY 288 CADVDGCSLAKEKTCVRKNENKCYNTPGSYVCVCPDGFEEFEDACVPAPAEATGSEPTOL 347  
 288 cadvdgcslakektcvrknencynptpgsyvcvcpdgdfeefedacvpapaeatgesptol 347

QY 288 cadvdgcslakektcvrknencynptpgsyvcvcpdgdfeefedacvpapaeatgesptol 347  
 288 cadvdgcslakektcvrknencynptpgsyvcvcpdgdfeefedacvpapaeatgesptol 347

QY 348 PSKREDL 353  
 348 pskredl 353

QY 348 pskredl 353  
 348 pskredl 353

RESULT 15  
 AAU31870  
 ID AAU31870 standard; Protein; 386 AA.  
 XX  
 AC AAU31870;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Novel human secreted protein #2361.  
 XX  
 KW Human; vaccination; gene therapy; nutritional supplement;  
 stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW PS Immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.

XX PN WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-0508656.

XX PR 18-APR-2000; 2000US-0552929.

XX PR 26-JAN-2001; 2001US-0770160.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-611725/70.

XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic

XX PT vaccination, testing and therapy -

XX PS Claim 20; Page 521; 765pp; English.

XX CC The invention relates to novel human secreted polypeptides. The

CC polypeptides and antibodies to the polypeptides are useful for

CC determining the presence of or predisposition to a disease associated

CC with altered levels of polypeptide. The polypeptides are also useful for

CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent

CC for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising

CC the nucleic acids encoding the polypeptides and cells genetically

CC engineered to express them are also useful for producing the proteins.

CC The proteins are useful in genetic vaccination, testing and

CC therapy, and can be used as nutritional supplements. They may be used to

CC increase stem cell proliferation; to regulate haematopoiesis; and in

CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;

CC immune suppression and/or stimulation; as anti-inflammatory agents; and

CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid

CC sequences of novel human secreted proteins of the invention.

XX SQ Sequence 386 AA;

Query Match 78.5%; Score 1573; DB 22; Length 386;  
 Best Local Similarity 74.9%; Pred. No. 5e-103;  
 Matches 286; Conservative 8; Mismatches 34; Indels 54; Gaps 5;

QY 21 APBAKKPFPCHRCRGLVNDKFNQGNVDTAKKNFGGNTAMEKTLSESSIRLLEIE 80  
 21 apbaakkpfpchrcrglvndkfngmntakknfggntameektlseessirllleie 80

QY 3 apeaakkpfpchrcrglvndkfngmntakknfggntameektlseessirllleie 62  
 3 apeaakkpfpchrcrglvndkfngmntakknfggntameektlseessirllleie 62

QY 81 GLCESSDFECNOMLEQNEHLEAMWLQKSEYFDLEFMCVTLVCCSPGTGTPCLOCG 140  
 81 glceessdfecnomleqnehleamwlqkseyfdlfevicvltlvccspgtgtptclcc 140

QY 63 glceessdfecnomleqnehleamwlqkseyfdlfevicvltlvccspgtgtptclcc 122  
 63 glceessdfecnomleqnehleamwlqkseyfdlfevicvltlvccspgtgtptclcc 122

QY 141 OCGSGRPSGNGHSGDSRSGDSRCHMGYGPPLCTDCMGYFS-SLRNETHSICT-- 197  
 141 ocgsgrpsgngshsgdsgsrgdsgsrchmgypplctdcmgyfs-slrnethsict-- 197

QY 123 gqsgsqgpcsgnghscgdsgrgdsgrchmgypplctdcmrllqlsrnethsictav 182  
 123 gqsgsqgpcsgnghscgdsgrgdsgrchmgypplctdcmrllqlsrnethsictav 182

QY 198 -----ACDESCRTCSGL 209  
 198 -----acdescrtcsagl 209

QY 183 rtglsdysppcclslgswrgvghawlrgrnltcpqysrrvwlaafspcdescckscgl 242  
 183 rtglsdysppcclslgswrgvghawlrgrnltcpqysrrvwlaafspcdescckscgl 242

QY 210 TNRDGECEVGNWLDGCAVDVDECAEPFCSAAGCKNANNGSYTCEECSSCVGCTGE 269  
 210 tnrdgecevgnwldgcavdvdecaepfcsaagcknannngsytyceecsscvgctge 269

QY 243 tnrdsgecevgwvrtvpaacqfcknpngsytyceecsscvgctge 302  
 243 tnrdsgecevgwvrtvpaacqfcknpngsytyceecsscvgctge 302

QY 270 GPNCCKECSISYAREHQCADVDEC-SLAKEKTCVRKNENKCYNTPGSYVCVCPDGFEE 327  
 270 gpncckecsisyarehqcadvdec-slakektcvrknencynptpgsyvcvcpdgdfe 327

QY 303 gpnckqclisyarehqcadvdecvphxpektlceektkcyntpgsyvcvcpdgdfe 362  
 303 gpnckqclisyarehqcadvdecvphxpektlceektkcyntpgsyvcvcpdgdfe 362



CC disease, pulmonary heart disease, rheumatic fever, congenital heart  
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,  
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,  
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,  
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune,  
 CC Albright syndrome, infertility, uterine disorders, viral disease. The  
 CC present sequence represents the human TANGO 206 polypeptide.

Sequence 420 AA:

Query Match 48.5%; Score 972.5; DB 22; Length 420;  
 Best Local Similarity 48.2%; Pred. No. 1e-60;  
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

OY 7 AALGLPLLLP-----PAEPA--KRPTRCRGLVDKFNQGVADAKKFFGG 56  
 DB 12 avlwglstflnlpplwlpqpsppspqphchtcrglyvstfngkylertldnfg 71  
 OY 57 NTAMEEKTLSKYESSEIRLEIEGLCESSDFECNOMLEAOEPHLEAWMLQKSEYDLE 116  
 DB 72 ntaweenlskylkdaetrlvelegvcsksdfetchnllalseelvwfhnqgeapdlf 131  
 OY 117 EMFCVKTLKVCSPGTGPDCLACGGSGRPGSGNCHSCSDSGRGGSCRMGYQPL 176  
 DB 132 gylcsdsklkcpgagtfpgscplpcpgyterpcgyqceqegtlrgsgnhdcaqygea 191  
 OY 177 CNDCHMDGYSSIRNETHSCTACDESKCTGNTNDCCECEVGNWLDGACVYDVECA 236  
 DB 192 cggcglgyfleaenashlvcsacifgpcarsgpeesncfckkqyvalhnlkcvdldcgt 251  
 OY 237 EPPPCSAQFCNANGSYTCECDSCVGTGEGPCNCKEISGYAREHGOCADVDECSL 296  
 DB 252 egangcadfcvntsgysycrcakacqcmagapgrckckspgyqyskoldvdece- 310  
 OY 297 AKRTCKRMKNCTNPGSTVCVCPDGFETEDACVP---PAEA---EATGE 342  
 DB 311 -levcpgenkqcenteggyrcicaeykqmeglcvkqjpesagffsentede 362

RESULT 18

AAV13362 standard; Protein; 420 AA.

AAV13362;

25-JUN-1999 (first entry)

Amino acid sequence of protein PRO214.

Secreted protein; transmembrane protein; human; enterocolitis;  
 Zollinger-Ellison syndrome; gastrointestinal ulceration;  
 congenital microvillus atrophy; skin disease; cell growth;  
 abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
 Parkinson's disease; Alzheimer's disease; AIDS; neuropathy;  
 fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
 anti-thrombotic; wound healing; tissue repair.

Homo sapiens.

MO9914328-A2.

25-MAR-1999.

16-SEP-1998; 98MO-US19330.

25-NOV-1997; 97US-0066840.

17-SEP-1997; 97US-0059113.

17-SEP-1997; 97US-0059115.

17-SEP-1997; 97US-0059117.

17-SEP-1997; 97US-0059119.

17-SEP-1997; 97US-0059121.

17-SEP-1997; 97US-0059122.  
 17-SEP-1997; 97US-0059184.  
 18-SEP-1997; 97US-0059263.  
 18-SEP-1997; 97US-0059266.  
 15-OCT-1997; 97US-0062125.  
 17-OCT-1997; 97US-0062285.  
 17-OCT-1997; 97US-0062287.  
 21-OCT-1997; 97US-0063486.  
 24-OCT-1997; 97US-0062814.  
 24-OCT-1997; 97US-0062816.  
 24-OCT-1997; 97US-0063045.  
 24-OCT-1997; 97US-0063120.  
 24-OCT-1997; 97US-0063121.  
 24-OCT-1997; 97US-0063127.  
 24-OCT-1997; 97US-0063128.  
 27-OCT-1997; 97US-0063329.  
 27-OCT-1997; 97US-0063327.  
 28-OCT-1997; 97US-0063541.  
 28-OCT-1997; 97US-0063542.  
 28-OCT-1997; 97US-0063544.  
 28-OCT-1997; 97US-0063549.  
 28-OCT-1997; 97US-0063550.  
 28-OCT-1997; 97US-0063564.  
 28-OCT-1997; 97US-0063435.  
 29-OCT-1997; 97US-0063704.  
 29-OCT-1997; 97US-0063732.  
 29-OCT-1997; 97US-0063738.  
 29-OCT-1997; 97US-0063734.  
 29-OCT-1997; 97US-0064215.  
 29-OCT-1997; 97US-0063735.  
 31-OCT-1997; 97US-0063870.  
 31-OCT-1997; 97US-0064103.  
 03-NOV-1997; 97US-0064248.  
 07-NOV-1997; 97US-0064809.  
 12-NOV-1997; 97US-0065186.  
 17-NOV-1997; 97US-0065846.  
 18-NOV-1997; 97US-0065893.  
 21-NOV-1997; 97US-0066120.  
 21-NOV-1997; 97US-0066364.  
 24-NOV-1997; 97US-0066772.  
 24-NOV-1997; 97US-0066466.  
 24-NOV-1997; 97US-0066770.  
 24-NOV-1997; 97US-0066511.  
 24-NOV-1997; 97US-0066453.

(GETH ) GENENTECH INC.

Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;

WPI; 1999-229533/19.

N-PSDB; AAX52233.

New isolated human genes and polypeptides used in, e.g. treatment of  
 gastrointestinal ulceration

Claim 12; Fig 40; 320pp; English.

AAV1344-403 represent secreted and transmembrane human proteins.  
 The cDNA sequences are obtained from cDNA libraries, prepared from  
 fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.  
 The encoded polypeptides have specific uses based on their homology to  
 known polypeptides, e.g. PRO211 and PRO212 can be used for disorders  
 associated with the preservation and maintenance of gastrointestinal  
 mucosa and the repair of acute and chronic mucosal lesions  
 (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal  
 ulceration and congenital microvillus atrophy), skin diseases associated  
 with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial  
 cancers such as lung squamous cell carcinoma of the vulva and gliomas),  
 potent effects on cell growth and development, diseases related to  
 growth or survival of nerve cells including Parkinson's disease,  
 Alzheimer's disease, AIDS, neuropathies or cancer. PRO265 can be used as  
 for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used  
 as a target for anti-tumor drugs. PRO533 may be used in the treatment

CC of Usner Syndrome or Atrophia areata. PRO269 can be used as an  
CC anti-thrombotic agent; PRO287 polypeptides and portions may have  
CC therapeutic applications in wound healing and tissue repair; PRO317 can  
CC be used for treating problems of the kidney, uterus, endometrium, blood  
CC vessels, or related tissue, e.g. in the heart of genital tract.

**SQ Sequence 420 AA;**

Query Match	48.5%	Score 971.5	DB 20	Length 420
Best Local Similarity	48.2%	Pred. No. 1.2e-60		
Matches 170	Conservative 48	Mismatches 116	Indels 19	Gaps 5

```

Oy 7 AALGILPILLPLP-----PAEEA--KKPCTCHRCRGIVDFENQOMDTAKKNGG 56
Db 12 avlwglsf flnlpwqwpqppqspqbphchcrglvdsfnkgletlrdfggg 71
Oy 57 NTAMBEKTLKSYESSSEIRLLEILEGLCBSSDFECNOMLAEQHEHLEAMWLOLKEYPDLF 118
Db 72 ntaweeenlskyksetrlvevlgevsksdfechrlllseteevlweswfmfkqeadlff 135
Oy 117 EMFCFKTLKVCSSGRTYGPDCLACOGGSGORSCNGHSCGDSHQSGSCCHNGYOGPL 176
Db 132 gwlsdsdsklccpgctfgrpsclprpbgtrterpcgyqgceggcttygsgnhdccagaygaa 191
Oy 177 CTDCMDGYFSSLRNETHSLTACDEBCKCTGSGLTNRDGCSECVGWLDEGACVDVDECAA 238
Db 192 cggcgslgyfteealnashlvcsacfgpcarcsgpseesnclqckkwaalhkvcdldecut 251
Oy 237 EPPPCSAQFCNMANGSYTCEECSSCGVCGEPPGMCKECTSGYAREHGCALVDECSL 296
Db 252 egangcgaqfvmtnegysyecrdaakaclgcmgagpryctckkpspyqvyqyscldvdece- 310
Oy 297 AEKTCVRKNENCMYTPGSYVCTPDGFEETDACY---PAEA---EATGE 342
Db 311 -levcpenkcenleeggyrcicegyqcmqglcvckeqipsaqlfsemtede 362

```

## RESULT 19

ID AAY05281 standard; Protein; 420 AA

AC AAY05281;

DT 22-JUN-1999 (first entry)

DE MEGE-like homologue PRO214.

KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;

KW EGF-8 homologue.

OS Homo sapiens.

PN 14327-A2

PD 25-MAR-1999.

PF 10-SEP-1998;

PF 10-SEP-1998; 98WO-US18824

PR 25-NOV-1997; 97US-0066840.

PR 17-SEP-1997; 97US-0059117.

PR	15-OCT-1997;	97US-0062125
PR	15-SEP-1997;	97US-0055203
PR	15-SEP-1997;	97US-0055203

PR 17-OCT-1997; 97US-0062287.

PR	29-OCT-1997;	97US-0063704
FN	24 OCT 1997;	97US-0002810

GENENTECH INC.

PI Botstein D, Goddard A, Gu

1

PI	Roy M, Wood WI;
XX	
DR	WPI; 1999-229532/19
DR	N-PSDB; AAX28431.
DR	

PT Antibodies against specific proteins overexpressed in tumours  
XX  
PS Example 1: Flg 10; 130pp; English.

CC This sequence represents the EGF-like homologue PRO214.  
CC The invention relates to antibodies (Ab) that bind to any of the  
CC polypeptides (1) designated PRO187; PRO533; PRO240; PRO211;  
CC PRO230; PRO261; PRO246 or EBAF-2. The Ab, or other agents that inhibit  
CC expression and/or activity of (1) are used: (1) to inhibit growth of  
CC tumours; and (11) as diagnostic/prognostic reagents for detection or  
CC quantification of (1) in cells or tissues, by standard immunoassays, with  
CC overexpression being indicative of cancer. For therapeutic use, the Ab  
CC may be conjugated to a toxin, chemotherapeutic agent or radioisotope.  
CC Genes expressing (1), many of which are growth factor homologues, are  
CC overexpressed in some cases of cancer.

sq Sequence 420 AA

Query Match	48.5%	Score	971.5	DB	20	Length	420
Best Local Similarity	48.2%	Pred. No.	1.2e-60				
Matches	170	Conservative	48	Mismatches	116	Indels	19
						Gaps	5

[illegible]

## RESULT 20

ID AAB24396 standard; Protein; 420 AA

... AAB24396

DT 07-NOV-2000 (first entry)

DE	Human PRO214	protein	sequence	SEQ ID NO:41
aa				

KM Human, PRO; promotion; inhibition; angiogenesis; cardiovascularisation;  
 KM diagnosis; trauma; wound; cancer; atherosclerosis; cardiac hypertrophy;  
 KM angiogenic; proliferative; candant; cardiovascular; antiatherosclerotic  
 KM cytostatic; gene therapy; vaccine.

OS Homo sapiens

PN WO200032221-A2

PD 08-JUN-2000



XX 30-NOV-1999; 99WO-US28313.  
 XX 01-DEC-1998; 98WO-US25108.  
 PR 16-DEC-1998; 98US-0112850.  
 PR 12-JAN-1999; 99US-0115554.  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 05-OCT-1999; 99WO-US23089.  
 PR 29-OCT-1999; 99US-0162506.

## (GETH ) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Hillan KJ, Goddard A;  
 PI Godowski PJ, Gurney AL, Klein RD, Kuo SS, Paoni NF, Smith V;  
 PI Watanabe CK, Williams PM, Wood WI;

WPI: 2000-412154/35.  
 N-PSDB; AAA77541.

XX Nucleic acids encoding PRO polypeptides useful for preventing,  
 PT diagnosing and treating diagnosing a cardiovascular, endothelial or  
 PT angiogenic disorders in mammals -  
 XX  
 XX Claim 72; Fig 18; 315pp; English.

XX The present invention describes nucleic acids encoding PRO polypeptides  
 CC useful for preventing, diagnosing and treating diagnosing a  
 CC cardiovascular, endothelial or angiogenic disorder in mammals by  
 CC modulating cell proliferation, angiogenesis and cardiovascularisation,  
 CC and for identifying agonists and antagonists of these processes. The  
 CC nucleic acids and the proteins they encode may be used in the  
 CC prevention, treatment and diagnosis of diseases associated with  
 CC inappropriate PRO expression such as cardiovascular, endothelial or  
 CC angiogenic disorders in mammals (e.g. atherosclerosis, cancers and  
 CC cardiac hypertrophy). For example, the nucleic acids (NCs) and vectors  
 CC containing them and the PRO polypeptide may be used to treat disorders  
 CC associated with decreased PRO expression. AAA77510 to AAA77721 and  
 CC AAB24388 to AAB24435 represent nucleotide and protein sequences used in  
 CC the exemplification of the present invention.  
 XX

SO Sequence 420 AA;

Query Match 48.5%; Score 971.5; DB 21; Length 420;  
 Best Local Similarity 48.2%; Pred. No. 1.2e-60;  
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

OY 7 AALGLPLLLLP-----PAPEAA--KKPTPCHRCRLVYKFNQGMWDTAKKNGG 56  
 DB 12 avtlwqstflnlpqplwlpqpsppqspqppqphchrcrglvsfknqglertlrdhfg99 71  
 OY 57 NPAWEKTLSSSESRILELLEGLCESSDFECNOMLEAOEHLPAWMLQKSEYDLP 116  
 DB 72 ntaweenlskydsetrlvrylgyvcsksdfchrllelseelvswwrthkxqgeapdlf 131  
 OY 117 EMEFCVTLAVCCSPGTGYGPDCLACGGSGORPCSGNGHCSGDGSRQDGSRCRHMGYQGPL 176  
 DB 132 qwlcsdsklkiccpagtfpcscplpcpgtferpgygcgeqgrfgrsgshdcdqaygsea 191  
 OY 177 CUDCMGCTSSSLRNETHSICLACDESKTCISGLTNRDCGCEVGVWLDGACVADVEBCA 236

DB 192 cggcglygfearnaahlyvsacifgpcarcsqpsencjgckkqwalhnlkcvdidecgt 251  
 OY 237 EPPPCSAAPCFCKNANGSYTCECDSSCVGCTGPGNCKCEICISGVARRHQCADVDECSL 296  
 DB 252 egancgadqfcvntlcsyeyrcdcaclcgmgagpprrckkcspgyqyqvskscldevece 310  
 OY 297 AEKTCVRKNECNYTPGSIYVCVCPDGFETEDACVP--PAEA---EATEGE 342  
 DB 311 -tevcpgenkqcentegyrclcaeykqmeqrlcvkxqipesagffsemtede 362

## RESULT 21

AA88569  
 ID AAY88569 standard; Protein; 420 AA.

XX AAY88569;

XX 09-AUG-2000 (first entry)

DE Human PRO214 amino acid sequence.

XX Antibody: PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246

KW PRO317; tumour growth inhibitor; cancer; diagnosis; treatment; human;

KW cell growth proliferation; HT protein; fibrulin; ADEPT;

XX antibody dependent enzyme mediated prodrug therapy.

OS Homo sapiens.

XX WO200015666-A2.

PD 23-MAR-2000.

PF 08-SEP-1999; 99WO-US20594.

PR 10-SEP-1998; 98US-0099803.

PR 10-SEP-1998; 98WO-US18824.

XX (GETH ) GENENTECH INC.

PI Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI, Botstein D;

DR WPI: 2000-271386/23.

XX N-PSDB; AAA30032.

XX New isolated antibodies which bind to specific polypeptides used for

PT diagnosis and treatment of neoplastic cell growth and proliferation -

XX Example 3; Fig 6; 200pp; English.

XX This sequence represents a human PRO214 amino acid sequence. PRO214

CC shares sequence homology with the HT protein and fibrulin. The

CC invention relates to isolated antibodies which bind to a polypeptide.

CC The "PRO" polypeptides are encoded by genes which are over expressed in

CC the genome of tumour cells. Vectors and host cells comprising the nucleic

CC acid encoding the antibodies are used in the production of the

CC antibodies. The antibodies and nucleic acids encoding them are used for

CC diagnosing a tumour in a mammal. The antibodies are used for inhibiting

CC the growth of tumour cells and identifying compounds that inhibit a

CC biological or immunological activity of and/or expression of a PRO187,

CC PRO533, PRO214, PRO240, PRO211, PRO230, PRO261, PRO246 or PRO317

CC polypeptide. The antibody can be used in antibody dependent enzyme

CC mediated prodrug therapy (ADEPT) by conjugating the antibody to a

CC prodrug-activating enzyme which converts a prodrug to an anti-cancer

CC drug. The antibodies can be fluorescently labelled and monitored by light

CC microscopy, flow cytometry or fluorometry for diagnosis and prognosis of

XX tumours.

SO Sequence 420 AA;

Query Match 48.5%; Score 971.5; DB 21; Length 420;  
 Best Local Similarity 48.2%; Pred. No. 1.2e-60;  
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;





PR 01-DEC-1999; 99WO-US28301.  
 PR 01-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28564.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 09-DEC-1999; 99WS-0170262.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US30911.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 PR 11-FEB-2000; 2000WO-US03565.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 22-FEB-2000; 2000WO-US04342.  
 PR 24-FEB-2000; 2000WO-US04914.  
 PR 24-FEB-2000; 2000WO-US05004.  
 PR 01-MAR-2000; 2000WO-US05601.  
 PR 20-MAR-2000; 2000WO-US07377.  
 PR 21-MAR-2000; 2000WO-US07532.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 10-NOV-2000; 2000WO-US30873.

(GETH ) GENENTECH INC.

XX Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W,  
 PI Gerltsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S,  
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
 DR N-PSDB: AAS21388.  
 WI: 2001-408281/43.

XX Isolated, secretory and transmembrane PRO polypeptide used to detect  
 PT other PRO polypeptides, link bioactive molecules to cells expressing  
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.  
 PT lung, breast, prostate, cervical  
 PS Claim 12; Fig 290; 813pp: English.

XX AAU12172-AAU12446 represent novel human secretory and transmembrane  
 CC PRO polypeptides. The PRO polypeptides are useful to detect other  
 CC PRO polypeptides, to link bioactive molecules to cells expressing  
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,  
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO  
 CC polypeptide expression in a cell sample to that in a control sample.  
 CC Some of the 275 sequences are also useful to stimulate the release of  
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the  
 CC proliferation or differentiation of chondrocytes, the proliferation or  
 CC gene expression in pericyte cells, the release of proteoglycans from  
 CC cartilage, the proliferation of inner ear utricular supporting cells or  
 CC of T-lymphocytes, the release of a cytokine from peripheral blood  
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of  
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by  
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide  
 CC to factor VITA. The PRO polypeptides can be used in assays to identify  
 CC molecules involved in binding interactions. The polynucleotides encoding  
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,  
 CC transgenic or knock out animals and can be used in gene therapy.

XX Sequence 420 AA:

Query Match 48.5%; Score 971.5; DB 22; Length 420;  
 Best Local Similarity 48.2%; Pred. No. 1.2e-60;  
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

QY 7 AALGLPLLLLP-----PAPEAA--KKPTPCRHCRGLVDKFNQGVATRAKKNFSGG 56

DB 12 avlwgslflnlpplwlpqppspqppqbpchctcrglvdstfngkterlctrdnfggg 71  
 QY 57 NTAMEEKLSTKYESSIRLEILEGICSSDFECNOMLEAOEHLEAMWLQKSEYDPLF 116  
 DB 72 ntaweeenlskydsctrlvevlyevcsksdfechllleseeiveswfhkqgeapdlf 131  
 QY 117 EMECVKTLKVCSPGTYPGDCIACOGSGRSCGNGHCSGDSROGDSCRCMHGOGPL 176  
 DB 132 gvlcsdsllkccpagtlfpsclpqpgrterpcgygqcegeglrgsgnhdccagyggea 191  
 QY 177 CTDCMDGFRSSLRNETHSICTACDESKTCGSLTRNDGCECEVGNWLDGACAVDDECA 236  
 DB 192 cgcqglgyfdaernashlyvcasclfpccarcsqpeesncilqckkyaalhklkcvdidecgt 251  
 QY 237 EPPPCSAOFCNNANSGYCECDSSVCCTGEGPCNCKECSGYAREHGOCADVDECSL 296  
 DB 252 egancgaqdgfcvntegsyecrdcaacalgcmagapbrckkcsppyyqyqgskcldvdece- 310  
 QY 297 AEKTCVRKNENCMYNTPGSYVCYCPDGFEEETEDACVP--PABA---EATEGE 342  
 DB 311 -fevcpngkngenteggyrcicaegykqmeglcvkeqlpesagffsentede 362

RESULT 24  
 AAB68594  
 ID AAB68594 standard; Protein; 420 AA.

XX AAB68594;

XX 27-APR-2001 (first entry)

XX PRO214.

XX Cytostatic; PRO protein; tumour; cancer.

XX Homo sapiens.

XX WO200105836-A1.

XX 25-JAN-2001.

XX 20-DEC-1999; 99WO-US30999.

XX 20-JUL-1999; 99US-0144758.

XX 08-SEP-1999; 99WO-US20594.

XX 13-SEP-1999; 99WO-US20944.

XX 15-SEP-1999; 99WO-US21090.

XX 05-OCT-1999; 99WO-US23089.

XX 29-NOV-1999; 99WO-US28214.

XX 30-NOV-1999; 99WO-US28313.

XX 02-DEC-1999; 99WO-US28564.

XX (GETH ) GENENTECH INC.

XX Botstein D, Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI;

XX WI: 2001-091968/10.

XX N-PSDB: AAF60352.

XX New antibody that binds to a PRO polypeptide, e.g. PRO187 and PRO533,

XX useful for diagnosing and treating cancers -

XX Claim 61; Fig 6; 196pp: English.  
 CC The present invention relates to PRO proteins and coding sequences. The  
 CC present sequence is one such PRO protein. It was found that the PRO genes  
 CC are amplified in the genome of tumour cells. The gene amplification is  
 CC expected to be associated with the overexpression of the gene product and  
 CC contributes to tumourigenesis. Therefore, antagonists of PRO proteins are  
 CC useful for the treatment of benign or malignant tumours, leukemias,  
 CC lymphoid malignancies and other disorders such as neuronal, glial,

CC astrocytal, hypothalamic, glandular, epithelial, inflammatory and  
 CC immunologic disorders.  
 XX  
 XX  
 SO Sequence 420 AA:

Query Match 48.5%; Score 971.5; DB 22; Length 420;  
 Best Local Similarity 48.2%; Pred. No. 1.2e-60;  
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

QY 7 AALGLPLLLLP-----PAPEAA--KKPPCHRCGLVDKFNQGVDTAKKNFGG 56  
 DB 12 avlwgslflnlpplwlpqpppspppphpchtcrglvdsfinkglertlrdfnfgg 71  
 QY NTAMEKRTLSKYSSERILLEGLCESDPEFCNQMLEAQBHEHLMWMLQKSEYDLE 116  
 DB 72 ntaweenlskydsctrlveleyvcsksdfechrlllseelvseswvthkqgeapdlf 131  
 QY 117 EWFVCVTLKVCSPGTYGPDCLACOGSORPCSGNGHSCGDSGROGDSGRCHMGYQPL 176  
 DB 132 qwlcsdsklkcpcagtfpsclpcpgtcrpggygcgegrtrgsyghdcqagyggea 191  
 QY 177 CTDCMDGYFSSLRNETHSICTACDESKTCGSLTRNDGCEYGVWLDGACVYDVECAA 236  
 DB 192 cggcglygfeaeernashlvcsacfcgpcarcspeesncclqckkqwalhnlkcvdidecgt 251  
 QY 237 EPPPCSAOFCRKNANGSYTCEDSSCGCTGRCGNCKECTISGYAREHQCADVDECSL 296  
 DB 252 egancgadqfcvnltegsyecdakacigcmagpgrckkcspgyqyqvsckldvdece- 310  
 QY 297 AEKTCVRKNENYNTPGSYVCVCPDGFEEEDACVP---PAEA---EATEGE 342  
 DB 311 -tevcpgeknqcenteggyrcicaegykkmeqglcvkeqldpesagffsemtede 362

RESULT 25  
 AAB80230  
 ID AAB80230 standard; Protein: 420 AA.  
 XX  
 AC AAB80230;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Human PRO214 protein.  
 XX  
 XX Human; PRO: dermatological; antipsoriatic; cytostatic; antinflammatory;  
 KW antiparkinsonian nootropic; neuroprotective; vulnerrary; cardiant;  
 KW antiangiogenic; vasotrophic; antiasthmatic; antihumalic; cancer;  
 KW antiarthritis; antifertility; antidiabetic; antilyral; diabetes;  
 KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;  
 KW Ischaemia; Inflammation.  
 XX  
 OS Homo sapiens.  
 PN WO200104311-A1.  
 XX  
 XX 18-JAN-2001.  
 PD  
 XX 22-FEB-2000; 2000WO-US04414.  
 PF  
 XX 07-JUL-1999; 99US-0143048.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 28-JUL-1999; 99US-0146222.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 13-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 05-OCT-1999; 99WO-US23089.  
 PR 29-NOV-1999; 99WO-US28214.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US30911.  
 PR 20-DEC-1999; 99WO-US30999.

PR 05-JAN-2000; 99WO-US00219.  
 XX  
 XX (GETH ) GENENTECH INC.  
 XX  
 PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;  
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;  
 PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavrin JI;  
 PI Mather JP, Pan Y, Paoni NF, Roy MA, Stewart TA, Tumas D;  
 PI Williams PM, Wood WT;  
 XX  
 XX WPI: 2001-081051/09.  
 DR N-PSDB; AAF72391.  
 DR  
 XX  
 PT Sixty one nucleic acids encoding PRO polypeptides which are useful in  
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung  
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.  
 PT Alzheimer's disease) -  
 XX  
 XX  
 PS Claim 1; Fig 40; 39pp; English.  
 CC  
 CC The present sequence is one of sixty one novel secreted and  
 CC transmembrane PRO polypeptides. The PRO polypeptides are  
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung  
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.  
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.  
 CC endometrial bleeding, angiodenesis, ischaemias such as coronary  
 CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,  
 CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and  
 CC diabetes and retinal disorders such as retinitis pigmentosum.  
 CC The PRO nucleic acids have applications in molecular biology, including  
 CC use as hybridization probes, and in chromosome and gene mapping.  
 XX  
 SQ Sequence 420 AA:

Query Match 48.5%; Score 971.5; DB 22; Length 420;  
 Best Local Similarity 48.2%; Pred. No. 1.2e-60;  
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

QY 7 AALGLPLLLLP-----PAPEAA--KKPPCHRCGLVDKFNQGVDTAKKNFGG 56  
 DB 12 avlwgslflnlpplwlpqpppspppphpchtcrglvdsfinkglertlrdfnfgg 71  
 QY 57 NTAMEKRTLSKYSSERILLEGLCESDPEFCNQMLEAQBHEHLMWMLQKSEYDLE 116  
 DB 72 ntaweenlskydsctrlveleyvcsksdfechrlllseelvseswvthkqgeapdlf 131  
 QY 117 EWFVCVTLKVCSPGTYGPDCLACOGSORPCSGNGHSCGDSGROGDSGRCHMGYQPL 176  
 DB 132 qwlcsdsklkcpcagtfpsclpcpgtcrpggygcgegrtrgsyghdcqagyggea 191  
 QY 192 cggcglygfeaeernashlvcsacfcgpcarcspeesncclqckkqwalhnlkcvdidecgt 251  
 QY 237 EPPPCSAOFCRKNANGSYTCEDSSCGCTGRCGNCKECTISGYAREHQCADVDECSL 296  
 DB 252 egancgadqfcvnltegsyecdakacigcmagpgrckkcspgyqyqvsckldvdece- 310  
 QY 297 AEKTCVRKNENYNTPGSYVCVCPDGFEEEDACVP---PAEA---EATEGE 342  
 DB 311 -tevcpgeknqcenteggyrcicaegykkmeqglcvkeqldpesagffsemtede 362

RESULT 26  
 AAB48133  
 ID AAB48133 standard; Protein: 420 AA.  
 XX  
 AC AAB48133;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX

DE	Human TANGO 206 variant 1 polypeptide.
XX	TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
KW	transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
KW	antithrombotic; neuroprotective; cytosolic; cardiant; hepatocytic;
KW	antiinflammatory; antidiabetic; antifertility; antipyletic; vasotropic;
KW	antirheumatic; nephrotropic; hemostatic; antileptic; osteopathic;
KW	ophthalmological; antisticking; antitumor; antitumor; variant.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Misc-difference 61
FT	/Label= E61D
FT	/note= "wild-type Glu is replaced by Asp"
PN	WO200069885-A2.
XX	
PD	23-NOV-2000.
XX	
PE	15-MAY-2000; 2000MO-US13361.
XX	
PR	14-MAY-1999; 99US-0312359.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Pan Y, Leiby KR;
XX	
DR	WPI; 2001-024999/03.
XX	
DR	N-PSDB; AHC84401.
XX	
PT	Novel nucleic acids encoding secreted or transmembrane proteins, useful
PT	for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
PT	the lung, liver, kidney or pancreas -
XX	
PS	Claim 8; Page -; 209pp; English.
XX	
CC	The invention provides human and mouse nucleic acids designated TANGO
CC	204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
CC	proteins. The polypeptides, nucleic acids and their modulators may be
CC	useful for treating or modulating cholesterol uptake, blood coagulation,
CC	to modulate cell proliferation, morphogenesis and fate specification,
CC	tissue repair and renewal, to treat cancer and promote wound healing,
CC	modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
CC	syndrome, protein S deficiency, modulate allergic or inflammatory
CC	response, acid secretion, tropic effects on gastrointestinal mucosa, and
CC	promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
CC	fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
CC	leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
CC	herniations, meningitis, ischemic brain or heart disease, infarction,
CC	intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
CC	disease, pulmonary heart disease, rheumatic fever, congenital heart
CC	disease, myocardial disease, atherosclerosis, hypertension, jaundice,
CC	hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC	sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC	disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
CC	neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC	Albright syndrome, infertility, uterine disorders, viral disease. The
CC	present sequence represents a human TANGO 206 variant polypeptide.
CC	Note: the present variant sequence has been constructed using the
CC	information provided in the specification.
XX	
50	Sequence 420 AA:

	Query Match	48.5%	Score 971.5;	DB 22;	Length 420;
	Best Local Similarity	48.2%;	Pred. No.	1.2e-60;	
	Matches 170; Conservative	48;	Mismatches	116;	Indels 19; Gaps 5;
Oy	7 AALGLPLLLLP-----PAPEAA-KRPTPCRGRGIVDKFNQGVADDAKKKEGGG	56			
	:                     :				
Dd	12 avlwgisiflnpgpiwqpsppqsppsqqphcctcrqlvdstinkgldrtlrdfnfygg	71			

57 NTAAEEKLSKYESESEIILLLELBSGLSSSPFECNOMLEAQBHLBAWMLDLKSEYDPLF 116  
 72 ntaweeenlsrykxdsdetrilveilegvcsksdfecfrlllletseeliveswffhkgqeaollf 131  
 117 EMFVCYKTLKVCSSPCTYTPDDCLACGCGSGRSCGNGHSCGSGSGRSGDSCRHNGYDPL 176  
 132 qwlcsdskllccpactlfpccpccgylccrpgygcgcgylrgysghcdqagaygaa 191  
 177 CTDCMDGYFSSLRNETHSICTACDESKTCSGLANRDCGECGVEGWLDEGACVDVDECAA 236  
 192 cggcgglgylfeeenashnlvcsacfcgpcarscspcesnclgckkqwalhhlkcvdldeqct 251  
 237 EPPFCSAAOFCKNANGSTYCEBDSSCYCGTGEFGPNKKECISGARHGCAVYDECSL 296  
 252 egangcagdqafcnlgeysyrcdrcaacacjgcmgagpyrckkpspyqgyskclvddece- 310  
 297 AEKTCVRNENCNYNPGSYVCVCPDGPFEPTDACP---PAEA----EATPGE 342  
 311 -tevcpengkqentleggyrclcaegykqmeglcvkqgipesaaffisemtde 362  
 RESULT 27  
 ID AAB27228  
 AAB27228 standard; Protein: 420 AA.  
 AAB27228;  
 27-MAR-2001 (first entry)  
 Human EXMAD-6 SEQ ID NO: 6.  
 Extracellular matrix and adhesion-associated protein; EXMAD; cancer;  
 inflammation; reproductive disorder; cardiovascular disorder;  
 immune disorder; musculoskeletal disorder; developmental disorder;  
 gastrointestinal disorder; cell proliferation disorder.  
 Homo sapiens.  
 OS  
 PN WO200068380-A2.  
 PD 16-NOV-2000.  
 PF 10-MAY-2000; 2000WO-US12811.  
 PR 11-MAY-1999; 99US-0133643.  
 PR 23-AUG-1999; 99US-0150409.  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DM;  
 PI Azimzai Y;  
 DR WPI: 2001-007395/01.  
 DR N-PSDB; AAC66895.  
 XX Isolated polynucleotide encoding extracellular matrix or  
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or  
 PT preventing disorders associated with expression of EXMAD such as  
 PT proliferative, immune and genetic disorders -  
 XX  
 Claim 1; Page 93-94; 129pp; English.  
 The present invention provides the protein and coding sequences for 25  
 CC novel extracellular matrix and adhesion-associated proteins (EXMADS).  
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,  
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,  
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,  
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are  
 CC useful in the prevention and treatment of cancers, cell proliferation,  
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and  
 CC gastrointestinal disorders and inflammation.  
 XX  
 Sequence 420 AA;  
 50

Query Match 48.5%; Score 971.5; DB 22; Length 420;  
Best Local Similarity 48.2%; Pred. NO. 1.2e-60;  
Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5.

RESULT	28
AAM41685	
ID	AAM41685 standard; Protein; 513 AA.

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6616.

KM Human;neurotopic; immunosuppressant; cytostatic; gene therapy; cancer  
KM peripheral nervous system; neuropathy; central nervous system; CNS;  
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KM chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KM leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR	21-JAN-2000	2000US-0488725
PR	25-APR-2000	2000US-0552317
PR	09-JUL-2000	2000US-0538042
PR	19-JUL-2000	2000US-06503410
PR	03-AUG-2000	2000US-0653450
PR	14-SEP-2000	2000US-0662191
PR	19-OCT-2000	2000US-0659306
PR	29-NOV-2000	2000US-0727334

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR WPI; 2001-442253/47.  
DR N-PSDB; AAI60841.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 2; SEQ ID NO 6616; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AA036642-AA042213) with neurotrophic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

50 Sequence 513 AA;

50 Sequence 513 AA;

Query Match	48.5%	Score 971.5	DB 22	Length 513
Best Local Similarity	48.2%	Pred. No. 1.5e-60		
Matches 170; Conservative	48;	Mismatches 116;	Indels 19;	Gaps 5

RESULT	29
AAB48134	
ID	AAB48134 standard; Protein; 420 AA

DT 02-APR-2001 (first entry)

Human TANGO 206 variant 2 polypeptide.

KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;  
 KW transmembrane protein; antiamebic; cerebroprotective; arteriosclerosis;  
 KW antiasthmatic; neuroprotective, cytoprotective; cardiatic; hepatotropic;  
 KW antiinflammatory; antidiabetic; antifertility; antipiretic; vasotropic;  
 KW antineumatic; nephroprotective; hemostatic; antilipemic; osteoprotic;  
 KW ophthalmological; antisticking; antilucer; vulnereary; variant.

OS Homo sapiens.

```

XX Key Location/Qualifiers
FH Misc-difference 76 /label= E76D
FT /note= "wild-type Glu is replaced by Asp"
XX
XX WO200069885-A2.
XX
XX 23-NOV-2000.
XX
XX 15-MAY-2000; 2000MO-US13361.
XX
XX 14-MAY-1999; 99US-0312359.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lei by KR;
XX
XX WPI: 2001-024999/03.
XX
XX N-PSDB: AAC84402.
XX
XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
XX the lung, liver, kidney or pancreas -
XX
XX Claim 8; Page -: 209pp; English.
XX
XX The invention provides human and mouse nucleic acids designated TANGO
XX CC TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
XX proteins. The polypeptides, nucleic acids and their modulators may be
XX useful for treating or modulating cholesterol uptake, blood coagulation,
XX to modulate cell proliferation, morphogenesis and fate specification,
XX tissue repair and renewal, to treat cancer and promote wound healing,
XX modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
XX syndrome, protein S deficiency, modulate allergic or inflammatory
XX response, acid secretion, tropic effects on gastrointestinal mucosa, and
XX promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
XX fibrous dysplasia, scoliosis, osteoarthritis, Goodpasture's syndrome,
XX leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
XX herniations, meningitis, ischemic brain or heart disease, infarction,
XX intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
XX disease, pulmonary heart disease, rheumatic fever, congenital heart
XX disease, myocardial disease, atherosclerosis, hypertension, jaundice,
XX hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
XX sickle cell disease, renal failure, ischemic bowel disease, Crohn's
XX disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
XX neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
XX Albright syndrome, infertility, uterine disorders, viral disease. The
XX present sequence represents a human TANGO 206 variant polypeptide.
XX Note: the present variant sequence has been constructed using the
XX information provided in the specification.
XX
XX Sequence 420 AA:

```

```

Query Match 48.4%; Score 969.5; DB 22; Length 420;
Best Local Similarity 47.9%; Pred. No. 1,6e-60;
Matches 169; Conservative 49; Mismatches 116; Indels 19; Gaps 5;

OY 7 AATGILLPLLLLP-----PAPAA--KKPTPCHRCRLVYKFNKGMDTAKKNGG 56
DB 12 avtawgslfllnlpqplwlgppppspqppchrcrlvdsfnkgldrtldhfg 71
OY 57 NTAMWEKTLKYSSEIRLEIEGLCESSDFPCNQMLAEQEHLEAMWLQKSEYDLP 116
DB 72 ntawdeenlskydserrlvelegvskdsfchrllelseelvwewfhkqgepdltf 131
OY 117 EMPFCVTKLVKCSBPGTYGPDCLACQGSQRCGSGNGHCSGDSGRQDGSRCRHMGYQGPL 176
DB 132 gtlcsdsklkicpcagtrtpscldpcpggterpcggygqcegegrtrgsghdcqagy9gea 191
OY 177 CTGCMGYSFSLRNHETICTACDESKTSGLTNRDGCCEGVWVLDACADVDECA 236
DB 177 CTGCMGYSFSLRNHETICTACDESKTSGLTNRDGCCEGVWVLDACADVDECA 236

```

```

DB 192 cggcg1gyfeaeernaahlyvcasacfpgarcsqpsesnclqckkyaallhllkcvdidecgt 251
OY 237 EPPPCSAAGFCCKNANGSYTCCECDSSCVGCTGEPGNCCKECISGYAREHGQCADVDECSL 296
DB 252 egangcgadqfcvntegsyecrcdakclcmagpprcckscppgyqgyskcdtdvdece- 310
OY 297 AERTCVKRNKENCYNTPGSYVCVCPDGFEEEDACVP---PAAE---EATBEB 342
DB 311 -tevcpngkqcenteggyrcloaeykqmgelcvkxqjpsaagffsente 362

RESULT 30
AAB48135
ID AAB48135 standard; Protein; 420 AA.
XX
XX AAB48135;
XX
XX 02-APR-2001 (first entry)
XX
XX Human TANGO 206 variant 3 polypeptide.
XX
XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
XX transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
XX antiasthmatic; neuroprotective; cytosstatic; cardiast; hepatotropic;
XX antiinflammatory; antidiabetic; antinfertility; antipyletic; vasotropic;
XX antirheumatic; nephrotropic; hemostatic; antilipemic; osteoplastic;
XX ophthalmological; antisickling; antilucer; vulnerary; variant.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 77 /label= E77D
FT /note= "wild-type Glu is replaced by Asp"
XX
XX WO200069885-A2.
XX
XX 23-NOV-2000.
XX
XX 15-MAY-2000; 2000MO-US13361.
XX
XX 14-MAY-1999; 99US-0312359.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lei by KR;
XX
XX WPI: 2001-024999/03.
XX
XX N-PSDB: AAC84403.
XX
XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
XX the lung, liver, kidney or pancreas -
XX
XX Claim 8; Page -: 209pp; English.
XX
XX The invention provides human and mouse nucleic acids designated TANGO
XX CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
XX proteins. The polypeptides, nucleic acids and their modulators may be
XX useful for treating or modulating cholesterol uptake, blood coagulation,
XX to modulate cell proliferation, morphogenesis and fate specification,
XX tissue repair and renewal, to treat cancer and promote wound healing,
XX modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
XX syndrome, protein S deficiency, modulate allergic or inflammatory
XX response, acid secretion, tropic effects on gastrointestinal mucosa, and
XX promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
XX fibrous dysplasia, scoliosis, osteoarthritis, Goodpasture's syndrome,
XX leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
XX herniations, meningitis, ischemic brain or heart disease, infarction,
XX intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
XX disease, pulmonary heart disease, rheumatic fever, congenital heart
XX disease, myocardial disease, atherosclerosis, hypertension, jaundice,
XX hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,

```

CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
 CC disease, hernias, hypoadrenalism, hyperadrenism, Cushing's syndrome,  
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McNamee  
 CC Albright syndrome, infertility, uterine disorders, viral disease. The  
 CC present sequence represents a human TANGO 206 variant polypeptide.  
 CC Note: the present variant sequence has been constructed using the  
 CC information provided in the specification.

XX Sequence 420 AA;

#### Query Match

Best Local Similarity 48.4%; Score 969.5; DB 22; Length 420;  
 Pred. No. 1.6e-60;  
 Matches 169; Conservative 49; Mismatches 116; Indels 19; Gaps 5;

QY 7 AALGLPLULLLP-----PAPEAA--KKPTCHRCRLVDFKNGAMDYAKKNGGG 56  
 Db 12 avlwgjlsflnlpqplwlpqpppsppqphpcrclrglvdslfnglertlrdnftg99 71  
 QY 57 NTAMEEKLTSKYESSSEIRLEILEGLCESSDFECNQMLEAOEHLLEAWMLQKSEYDLE 116  
 Db 72 ntawedenlskykdserrlvelegvcsksdfechrllleseeelweswflkqgeapdlf 131  
 QY 117 EMFCVKTLLKVCSPGTYGPDCLACQGSQRPSCGNGHCSGDSGSGRCHMGYOGPL 176  
 Db 132 qwlcsdsklkcpagltfpcpqlpcpglterpcggygqcegegrtrgsghdcqagysgea 191  
 QY 177 CTDCMGYFSSLRNETHSICACDESKTCSGLTNRCGCEVGVWVDEGACVYDCCAA 236  
 Db 192 cggcgjlgylfeearnashlvcsacftgpcarcspspeesncldckkqywalhhlkcvdidecgt 251  
 QY 237 EPPPSAAGFCCKNANGSYTCCECDSSCVGCTGEGPGNCKECISGYAREHGOACADVDCSL 296  
 Db 252 egancgaddqfvcntegsyeccidckacldgcmagaprcckkspgyqgyskclvddece- 310  
 QY 297 AEKTCVRKNENCYNTPGSYVCVCPDGFETEDACVP--PAEA---EATGE 342  
 Db 311 -tevcpgekqcenteggyrcicaeygkqmeqlcvkeqldesagffseamlede 362

#### RESULT 31

AAB38395 AAB38395 standard; Protein; 434 AA.

AC AAB38395;  
 DT 31-JAN-2001 (first entry)

Human secreted protein encoded by gene 56 clone HD7AT90.

XX Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;  
 KM cytostatic; cardiant; vasotrophic; cerebroprotective; neuroprotective;  
 KM neotropic; antibacterial; vitucide; fungicide; ophthalmological; human;  
 KM vulnerary; gene therapy; infection; secreted protein.

OS Homo sapiens.

PN MO2000061623-A1.

PD 19-OCT-2000.

PF 06-APR-2000; 2000MO-US08979.

PR 09-APR-1999; 99US-0128693.

PR 26-APR-1999; 99US-0130991.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Ni J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;  
 PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;  
 PI Young PE;

DR WPI; 2000-647418/62.

XX New nucleic acid molecules encoding 62 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -

PS Claim 11; Page 652-653; 716pp; English.

XX Sequences AAB38321-B38396 represent the amino acid sequences of 62  
 CC human secreted proteins encoded by the genes AAC69512-C69587. The genes  
 CC and proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
 CC a range of human tissues disclosed in the specification. The nucleic  
 CC acids, proteins, antibodies and (ant)agonists are useful in the  
 CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.  
 CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms  
 CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac  
 CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)  
 CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)  
 CC infections caused by bacteria, viruses and fungi; and (h) ocular  
 CC disorders e.g. corneal infection. The polypeptides can also be used to  
 CC aid wound healing and epithelial cell proliferation, to prevent skin  
 CC aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis.

XX Sequence 434 AA;

#### Query Match

Best Local Similarity 48.3%; Score 969; DB 21; Length 434;  
 Pred. No. 1.8e-60;  
 Matches 168; Conservative 50; Mismatches 119; Indels 20; Gaps 4;

QY 7 AALGLPLULLLP-----PAPEAA--KKPTCHRCRLVDFKNGAMDYAKKNGGG 56  
 Db 9 avlwgjlsflnlpqplwlpqpppsppqphpcrclrglvdslfnglertlrdnftg99 68  
 QY 57 NTAMEEKLTSKYESSSEIRLEILEGLCESSDFECNQMLEAOEHLLEAWMLQKSEYDLE 116  
 Db 69 ntawedenlskykdserrlvelegvcsksdfechrllleseeelweswflkqgeapdlf 128  
 QY 117 EMFCVKTLLKVCSPGTYGPDCLACQGSQRPSCGNGHCSGDSGSGRCHMGYOGPL 176  
 Db 129 qwlcsdsklkcpagltfpcpqlpcpglterpcggygqcegegrtrgsghdcqagysgea 188  
 QY 177 CTDCMGYFSSLRNETHSICACDESKTCSGLTNRCGCEVGVWVDEGACVYDCCAA 236  
 Db 189 cggcgjlgylfeearnashlvcsacftgpcarcspspeesncldckkqywalhhlkcvdidecgt 248  
 QY 237 EPPPSAAGFCCKNANGSYTCCECDSSCVGCTGEGPGNCKECISGYAREHGOACADVDCSL 296  
 Db 249 egancgaddqfvcntegsyeccidckacldgcmagaprcckkspgyqgyskclvddece- 307  
 QY 297 AEKTCVRKNENCYNTPGSYVCVCPDGFETEDACVPAEAATEGESPTQPSREDL 353  
 Db 308 -tevcpgekqcenteggyrcicaeygkqmeqlcv-----kqldpafpdltdl 355

#### RESULT 32

AAV76151 AAV76151 standard; Protein; 434 AA.

AC AAV76151;

DT 23-MAR-2000 (first entry)

Human secreted protein encoded by gene 28.

XX Human; secreted protein; cancer; tumour; developmental abnormality;  
 KM foetal deficiency; blood disorder; immune system disorder; inflammation;  
 KM autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
 KM schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;  
 KM atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
 KM digestive disorder; endocrine disorder; infection; AIDS; leukaemia;



therapy; chromosome 3.

XX OS Homo sapiens.  
XX PN M09958660-A1  
XX PD 18-NOV-1999.

XX PF 06-MAY-1999; 99WO-US09847.  
XX PR 12-MAY-1998; 98US-0085093.  
PR 12-MAY-1998; 98US-0085094.  
PR 12-MAY-1998; 98US-0085105.  
PR 12-MAY-1998; 98US-0085180.  
PR 18-MAY-1998; 98US-0085906.  
PR 18-MAY-1998; 98US-0085920.  
PR 18-MAY-1998; 98US-0085921.  
PR 18-MAY-1998; 98US-0085922.  
PR 18-MAY-1998; 98US-0085923.  
PR 18-MAY-1998; 98US-0085924.  
PR 18-MAY-1998; 98US-0085928.  
PR 18-MAY-1998; 98US-0085925.  
PR 18-MAY-1998; 98US-0085927.

XX PA (HUMAN) HUMAN GENOME SCI. INC.  
XX PI RUBEN SM, Florence K, Ni J, Rosen CA, Carter KC, Moore PA;  
PI Olsen HS, Shi Y, Young PE, Wei F, Brewer LA, Soppet DR;  
PI Lafleur DW, Endress GA, Ebner R;  
XX WPI: 2000-062296/05.  
XX N-PSDB: AA65277.

XX PT New isolated human genes and the secreted polypeptides they encode,  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders -  
XX

XX PS Claim 11; Page 380-381; 475pp; English.

XX AA65250 to AA65350 represent 97 isolated human secreted protein genes.  
CC AA65250 to AA65350 are the secreted proteins encoded by the 97 human  
CC genes. The gene encoding this protein was found to be on chromosome 3.  
CC The genes and their corresponding secreted polypeptides are  
CC useful for preventing, treating or ameliorating medical conditions,  
CC e.g. by protein or gene therapy. Also pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new genes. Specific  
CC uses are described for each of the 97 genes, based on which tissues they  
CC are most highly expressed in, and include developing products for the  
CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
CC and foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin  
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
CC disorders, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The sequences shown in AA65224 to AA65424 represent fragments of the  
CC secreted proteins.  
XX

XX SO Sequence 434 AA:

Query Match 48.3%; Score 969; DB 21; Length 434;  
Best Local Similarity 47.1%; Pred. No. 1,8e-60;  
Matches 168; Conservative 50; Mismatches 119; Indels 20; Gaps 4;

OY 7 AALGLPLLLLP-----PAPEAA--KKPPCHRCGLVDKENGQAVDRAKKKFGCG 56  
DB 9 AYLWGLSLINLPYPIWLPSP 68  
OY 57 NPAWEKTSIKYESSIRLEILEIGCESDFPCNOMLEQOEHLAWMLQKSEYPDF 116  
DB 69 NPAWEENLSKYKDSRIIVLVGVCSKSDFECHLILSEELVESWVFNHKGQEADLI 128

OY 117 EMFCVTKLRVCCSPGTYPGDCLACGGSGRPGCSGNCHCSGDSRGDSCRCRHMVGQPL 176  
DB 129 GWLCSDSIKLCCPAGTfGpScLpcpggterpcgy9qcegegrtgrsgnccdcagyygga 188  
OY 177 CTDCMDGVRSSLRNETHSHCTACDESKCTCSGLTRNDGCEVGVWLDGACVVDDECA 236  
DB 189 CGCGJLYfEaerNashVCSaCfGpcarCSgPeeSnCfGCKKyaLhKcVdIdEcgt 248  
OY 237 EPPPCSAAPFCRKNANGSYTCCECDSSCVGCTGEGPNCCKECISGYAREHGOCADDECSTL 296  
DB 249 EGANGCAGDfCvntGhsyCircdCaKacLgcmAgpgrCKcspYqYgysKcLdVdEcE- 307  
OY 297 AEKTCVRKNENCYNTPGSYVCVCPDGFETEDACYPPAAEATEGSPQLPSRDL 353  
DB 308 -levcpGnkGcEntegYrCfCaegYkqmeGfCv-----keqJpafalcdl 355

RESULT 33  
AAB48110  
ID AAB48110 standard; Protein; 420 AA.  
XX  
AC AAB48110;  
XX  
DT 02-APR-2001 (first entry)  
XX  
XX Mouse TANGO 206 polypeptide.  
DE  
XX TANGO 204; TANGO 206; A236; secreted protein; human; mouse;  
KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;  
KW antiasthmatic; neuroprotective; cyostatic; cardiant; hepatotropic;  
KW antiinflammatory; antidiabetic; antinfertility; antipyletic; vasotropic;  
KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;  
KW ophthalmological; antisticking; antilucer; vulnary.  
XX  
OS Mus sp.  
XX  
XX WO200069885-A2.  
XX PN  
XX PD 23-NOV-2000.  
XX PR 15-MAY-2000; 2000WO-US13361.  
XX PR 14-MAY-1999; 99US-0312359.  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX PI Pan Y, Leiby KR;  
XX WPI: 2001-024999/03.  
XX N-PSDB: AAC84387; AAC84388.  
DR  
XX Novel nucleic acids encoding secreted or transmembrane proteins, useful  
PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of  
PT the lung, liver, kidney or pancreas -  
XX  
XX PS Claim 8; Fig 11A-D; 209pp; English.

CC The invention provides human and mouse nucleic acids designated TANGO  
CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane  
CC proteins. The polypeptides, nucleic acids and their modulators may be  
CC useful for treating or modulating cholesterol uptake, blood coagulation,  
CC tissue repair and renewal, to treat cancer and promote wound healing,  
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
CC syndrome, protein S deficiency, modulate allergic or inflammatory  
CC response, acid secretion, tropic effects on gastrointestinal mucosa, and  
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,  
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis;  
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain  
CC herniations, meningitis, ischemic brain or heart disease, infarction,  
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart  
CC disease, pulmonary heart disease, rheumatic fever, congenital heart



CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,  
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,  
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,  
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune  
CC Albright syndrome, infertility, uterine disorders, viral disease. The  
CC present sequence represents the mouse TANGO 206 polypeptide.

SO Sequence 420 AA;

Query Match 48.3%; Score 968.5; DB 22; Length 420;  
Best Local Similarity 47.0%; Pred. No. 1.9e-60;  
Matches 171; Conservative 43; Mismatches 121; Indels 29; Gaps 5;

QY 3 LPRRAALGLPLULLL-----PPAPEAKKPPCHRCGLVKNFGM 45  
DB 4 lprp---glvpsllwclslflslpqpwlqpspphspiraephchrcrlalvnhfkgyl 60  
QY 46 VDTAKKNGGNTAMEEKTLSKYSSEIRLEILEGLCESSDFPCNOMLEAOEHLLEAMW 105  
DB 61 ertlrndfgygnntaweekiskykdselrivelvgyvcsrdsfchrlleseelevenw 120  
QY 106 LQLKSEYVDLFEWPCVTKLVKCCSPGTYPDCLACOGSGORPCSGNGSCSDGSGNOGDS 165  
DB 121 fhrgqeadplfqwlcscslklccpsgtfcpqgtlpcpgtlerpcgygqcegegrtgsqh 180  
QY 166 CRCHMGVQGPLCTDCMNGYSSLRNETHSICACDESCRTSGLTNRDCCGEVGVWIDE 225  
DB 181 cdcgagygagcgcgqglgyfcaernshlvcsacifgpcarctgpeesbhcldckkqywlh 240  
QY 226 GACVDVDECAAEPPPCSAOFCCKNANSYTCEDSSCVGCTGSGPGCKCKICISGVAREH 285  
DB 241 lkcvdidecgteqatcgadqfcvntcegsyecrdcaakclgcmgagpgrckcsrgyqvg 300  
QY 286 GQCADVDECSLAETCYRKKNMCYNTPGSYVCVPDGFETEDAC---VPPAE---AEA 338  
DB 301 skcldvdecetv--vcpgenekcenteggyrcvcaegyrgedglcvkqepesagffae 358  
QY 339 TEGE 342  
DB 359 tede 362

RESULT 34

AAB48137 standard; Protein: 420 AA.

XX AAB48137;

DT 02-APR-2001 (first entry)

DE Mouse TANGO 206 variant 2 polypeptide.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;  
XX transmembrane protein; anti-neuritic; cerebroprotective; arteriosclerosis;  
XX anti-inflammatory; neuroprotective; cytoskeletal; cardiatic; hepatotropic;  
XX anti-neuritic; anti-diabetic; anti-inflammation; anti-pyretic; vasotropic;  
XX anti-neuritic; nephrotropic; hemostatic; anti-lipemic; osteopathic;  
XX ophthalmological; antisticking; anti-ulcer; vulnerary; variant.

OS Mus sp.

XX Key Location/Qualifiers

FT Misc-difference 61

FT /label= B61D

XX /note= "wild-type glu is replaced by asp"

XX MO200069885-A2.

XX 23-NOV-2000.

XX 15-MAY-2000; 2000MO-US13361.

XX 14-MAY-1999; 99US-0312359.  
XX (MILL-) MILLENNIUM PHARM INC.  
XX Pan Y, Leiby KR;

XX WPI: 2001-024999/03.  
XX N-PSDB: AAC84405.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful  
XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of  
XX the lung, liver, kidney or pancreas -  
XX Claim 8; Page -: 209pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO  
XX 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane  
XX proteins. The polypeptides, nucleic acids and their modulators may be  
XX useful for treating or modulating cholesterol uptake, blood coagulation,  
XX to modulate cell proliferation, morphogenesis and fate specification,  
XX tissue repair and renewal, to treat cancer and promote wound healing,  
XX modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
XX syndrome, protein S deficiency, modulate allergic or inflammatory  
XX response, acid secretion, tropic effects on gastrointestinal mucosa, and  
XX promote ulcer healing, treat bone cancer, achondroplasia, myeloma,  
XX fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,  
XX leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain  
XX herniations, meningitis, ischemic brain or heart disease, infarction,  
XX intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart  
XX disease, pulmonary heart disease, rheumatic fever, congenital heart  
XX disease, myocardial disease, atherosclerosis, hypertension, jaundice,  
XX hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,  
XX sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
XX disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,  
XX neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune  
XX Albright syndrome, infertility, uterine disorders, viral disease. The  
XX present sequence represents a mouse TANGO 206 variant polypeptide.  
XX Note: the present variant sequence has been constructed using the  
XX information provided in the specification.

SO Sequence 420 AA;

Query Match 48.3%; Score 967.5; DB 22; Length 420;  
Best Local Similarity 47.0%; Pred. No. 2.3e-60;  
Matches 171; Conservative 43; Mismatches 121; Indels 29; Gaps 5;

QY 3 LPRRAALGLPLULLL-----PPAPEAKKPPCHRCGLVKNFGM 45  
DB 4 lprp---glvpsllwclslflslpqpwlqpspphspiraephchrcrlalvnhfkgyl 60  
QY 46 VDTAKKNGGNTAMEEKTLSKYSSEIRLEILEGLCESSDFPCNOMLEAOEHLLEAMW 105  
DB 61 ertlrndfgygnntaweekiskykdselrivelvgyvcsrdsfchrlleseelevenw 120  
QY 106 LQLKSEYVDLFEWPCVTKLVKCCSPGTYPDCLACOGSGORPCSGNGSCSDGSGNOGDS 165  
DB 121 fhrgqeadplfqwlcscslklccpsgtfcpqgtlpcpgtlerpcgygqcegegrtgsqh 180  
QY 166 CRCHMGVQGPLCTDCMNGYSSLRNETHSICACDESCRTSGLTNRDCCGEVGVWIDE 225  
DB 181 cdcgagygagcgcgqglgyfcaernshlvcsacifgpcarctgpeesbhcldckkqywlh 240  
QY 226 GACVDVDECAAEPPPCSAOFCCKNANSYTCEDSSCVGCTGSGPGCKCKICISGVAREH 285  
DB 241 lkcvdidecgteqatcgadqfcvntcegsyecrdcaakclgcmgagpgrckcsrgyqvg 300  
QY 286 GQCADVDECSLAETCYRKKNMCYNTPGSYVCVPDGFETEDAC---VPPAE---AEA 338  
DB 301 skcldvdecetv--vcpgenekcenteggyrcvcaegyrgedglcvkqepesagffae 358  
QY 339 TEGE 342

Dn	359 tede 362
<hr/>	
RESULT	35
ID	AAB48136
AC	AAB48136 standard; Protein; 420 AA.
XX	
AA	AAB48136;
DT	02-APR-2001 (first entry)
DE	
XX	
XX	Mouse TANGO 206 variant 1 polypeptide.
KM	TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
KW	transmembrane protein; antineuritic; cerebroprotective; arteriosclerosis;
KW	antiaesthetic; neuroprotective, cytosolic; cardiac; hepatotropic;
KW	antiinflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;
KW	antirheumatic; nephrotoxic; hemostatic; antileptic; osteopathic;
KW	ophthalmological; antisticking; antifencer; vulnerary; variant.
XX	
OS	Mus sp.
FT	
Key	Location/Qualifiers
FT	Misc-difference 42 /label= E42D
FT	/note="wild-type Glu is replaced by Asp"
XX	
PN	MOZ00069885-A2.
PD	
PD	23-NOV-2000.
PF	
PF	15-MAY-2000; 2000WO-US13361.
PR	
PR	14-MAY-1999; 99US-0312359.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
PI	
PI	Pan Y, Leiby KR;
DR	WPI: 2001-024999/03.
DR	N-PSDB: AAC84404.
PT	
PT	Novel nucleic acids encoding secreted or transmembrane proteins, useful
PT	for treating, e.g., cancer, hemophilia, anemia, ischemia or diseases of
PT	the lung, liver, kidney or pancreas -
XX	
PS	
PS	Claim 8; Page - : 209pp; English.
CC	
CC	The invention provides human and mouse nucleic acids designated TANGO
CC	204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
CC	proteins. The polypeptides, nucleic acids and their modulators may be
CC	useful for treating or modulating cholesterol uptake, blood coagulation,
CC	to modulate cell proliferation, morphogenesis and fate specification,
CC	tissue repair and renewal, to treat cancer and promote wound healing,
CC	modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
CC	syndrome, protein S deficiency, modulate allergic or inflammatory
CC	response, acid secretion, tropic effects on gastrointestinal mucosa, and
CC	promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
CC	Crohn's dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis
CC	leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
CC	herniations, meningitis, ischemic brain or heart disease, infarction,
CC	intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
CC	disease, pulmonary heart disease, rheumatic fever, congenital heart
CC	disease, myocardial disease, atherosclerosis, hypertension, jaundice,
CC	hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC	sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC	disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
CC	neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC	Albright syndrome, infertility, uterine disorders, viral disease. The
CC	present sequence represents a mouse TANGO 206 variant polypeptide.
CC	Note: The present variant sequence has been constructed using the
CC	information provided in the specification.

XX	Sequence	420 AA;
SO	Query Match	48.2%, Score 966.5; DB 22; Length 420;
	Best Local Similarity	47.0%, Pred. No. 2.7e-60;
	Matches 171; Conservative	42; Mismatches 122; Indels 29; Gaps 5;
OY	3 LPRRAAGLILPLULLL-----PPAPFAKKPRPCHRCRGLVDKFNQGM 45	
Db	4 LPR-----GLVSLILWCLSLFISLPGPVWLPQSPHPSPRADPHCHCRALVDNFKY 60	
OY	46 VDTAKKNFGGNTAMWEEKTLKSYESSSEIRLLEILGLCESSDFECNQMLEAOEHLKAW 105	
Db	61 erlrldnfggnntaweeklaskydserrlveylagvcsrdsfchrllelseelvenmw 120	
OY	106 LQKSEYDLEFEMFVKTLKVCSSFGRTGPPCLACQGSQSRPCSGNHCSSDGSNQGGS 165	
Db	121 fhrggeadellfwlcsdslikiccpqgtltpesclpcpgtlerpcpgyggqcegeqlrfgsgh 180	
OY	166 CRCHNGYGPPLCTDCMDGYFSLRNETHSICACDESKTSGLTNRRCGCEGVWVDE 225	
Db	181 cdccgagyggeacqcgqlyfeamsmshlvesactgparcrtgpesnclqckkwalh 240	
OY	226 GACVDVDECAAPPPCSAAQPCKNANSGYTCHECDSSCVGCTGBEGPKCKECISYAREH 285	
Db	241 lkcvdidecgeqatcgadqfcmvtegsyercdcakacilgcmagpgrckkcsryggvq 300	
OY	286 GGCADVDECSLAIEKTCVKKKNENCYVTPGSYVCVCPDGFEEFEDAC---VPPAE---AEA 338	
Db	301 skclvdveceev--vcpgenckenteggyrcvcaegyrgegdilcvkqyvesagfiadm 358	
OY	339 TEGE 342	
Db	359 tede 362	
RESULT 36		
AAB48138		
ID	AAB48138 standard; Protein: 420 AA.	
AC	AAB48138;	
XX	02-APR-2001 (first entry)	
DE	Mouse TANGO 206 variant 3 polypeptide.	
XX	TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;	
KW	transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;	
KW	antiallergic; neuroprotective; cytosolic; candid; hepatotropic;	
KW	antiinflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;	
KW	antirheumatic; nephrotropic; hemostatic; antileptic; osteopathic;	
KW	ophthalmological; antislaking; antiulcer; vulnary; variant.	
XX	Mus sp.	
XX	Key	Location/Qualifiers
EH	Misc-difference 76	/label= E76D
FT	/note= "wild-type glu is replaced by asp"	
XX	MO200069885-A2.	
XX	23-NOV-2000.	
XX	15-MAY-2000; 2000WO-US13361.	
XX	14-MAY-1999; 99US-0312359.	
XX	(MILL-) MILLENNIUM PHARM INC.	
XX	Pan Y, Leiby KR.	

DR WPI: 2001-024999/03.  
 DR N-PSDB: AAC84406.  
 XX  
 PT Novel nucleic acids encoding secreted or transmembrane proteins, useful  
 PT for treating, e.g., cancer, hemophilia, anemia, ischemia or diseases of  
 PT the lung, liver, kidney or pancreas -  
 XX  
 PS Claim 8; Page -: 209pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO  
 CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane  
 CC proteins. The polypeptides, nucleic acids and their modulators may be  
 CC useful for treating or modulating cholesterol uptake, blood coagulation,  
 CC to modulate cell proliferation, morphogenesis and fate specification,  
 CC tissue repair and renewal, to treat cancer and promote wound healing,  
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
 CC syndrome, protein S deficiency, modulate allergic or inflammatory  
 CC response, acid secretion, trophic effects on gastrointestinal mucosa, and  
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,  
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,  
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain  
 CC herniations, meningitis, ischemic brain or heart disease, infarction,  
 CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart  
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart  
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,  
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,  
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,  
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune  
 CC Albright syndrome, infertility, uterine disorders, viral disease. The  
 CC present sequence represents a mouse TANGO 206 variant polypeptide.  
 CC Note: the present variant sequence has been constructed using the  
 CC information provided in the specification.  
 CC  
 XX  
 SQ Sequence 420 AA;

Query Match 48.2%; Score 965.5; DB 22; Length 420;  
 Best Local Similarity 46.7%; Pred. No. 3.1e-60;  
 Matches 170; Conservative 44; Mismatches 121; Indels 29; Gaps 5;

QY 3 LPRAALGLPLLL-----PPAPAAKKPPPHRCGLVDKFNQGM 45  
 Db 4 LPPL--GLVPLIWCISLFLSLPGVWLQPSPPHPSPREHPCHCRALVDINFKGI 60  
 QY 46 VDTAKKNGGNTAMEKTLKSYSESEIRLEILEGLCESSDFECNOMLEAOEHLAMW 105  
 Db 61 ertlrdnfggntawdeeklskysdetrlyevlegvstrsdfechrllseelvenw 120  
 QY 106 LQAKSEYPDLEFPCVKTLYKCCSPGTYPDCLACQCGSQRCGSGNGHSGDGRQDGS 165  
 Db 121 fhqgeapdlftqalcslkicpsgtfpcplpcpggtterpcggygcegegrtgrgsh 180  
 QY 166 CRCHMGVQGLPCTDCMDCGYFSSLNETHSTCTACDESKTSGLTNRCCGECENGVWIDE 225  
 Db 181 cdcaagyyggaagcgqjlyfleaernshlvcsactfgycparctgpeeshclqkkywaih 240  
 QY 226 GACVDVDECAAEPPPCSAOFCCKNANGSYTCEDDSCVGTGEGSPGCKICISYAREH 285  
 Db 241 lkcvdidecgeatcgadqgfcvnlcegsyecrdcaklqcmgaqpgtckcsryqgvg 300  
 QY 286 GQCADVDECSLAERTCVKRNKNCYNTPGSYVVCVCPDGFEEEDAC---VPPAE---AEA 338  
 Db 301 skclvdcectev--vcpgekenkcenteggyrcvcaegyrtdgdglcvkqepesagffem 358  
 QY 339 TEGE 342  
 Db 359 tede 362

RESULT 37  
 AAB38394  
 ID AAB38394 standard; Protein: 392 AA.

XX  
 AC AAB38394;  
 XX

XX 31-JAN-2001 (first entry)

XX Human secreted protein encoded by gene 56 clone HPRAL78.

XX Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;  
 KW cytostatic; cardiant; vasotropic; cerebroprotective; neuroprotective;  
 KW nototropic; antibacterial; virucide; fungicide; ophthalmological; human;  
 KW vulnertary; gene therapy; infection; secreted protein.

XX Homo sapiens.

XX WO200061623-A1.

XX 19-OCT-2000.

XX 06-APR-2000; 2000WO-US08979.

XX 09-APR-1999; 99US-0128693.

XX 26-APR-1999; 99US-0130991.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, N1 J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;  
 PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;  
 PI Young PE;

XX WPI: 2000-647418/62.

XX New nucleic acid molecules encoding 62 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 XX

XX Claim 11; Page 650-651; 716pp; English.

XX Sequences AAB38321-B38396 represent the amino acid sequences of 62  
 CC human secreted proteins encoded by the genes AAC69512-C69587. The genes  
 CC and proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g., by protein or gene therapy. The genes are isolated from  
 CC a range of human tissues disclosed in the specification. The nucleic  
 CC acids, proteins, antibodies and (ant)agonists are useful in the  
 CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.,  
 CC rheumatoid arthritis; (b) hyperproliferative disorders e.g., neoplasms  
 CC of the breast or liver; (c) cardiovascular disorders e.g., cardiac  
 CC arrest; (d) cerebrovascular disorders e.g., cerebral ischemia; (e)  
 CC angiogenesis; (f) nervous system disorders e.g., Alzheimer's disease; (g)  
 CC infections caused by bacteria, viruses and fungi; and (h) ocular  
 CC disorders e.g., corneal infection. The polypeptides can also be used to  
 CC aid wound healing and epithelial cell proliferation, to prevent skin  
 CC aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis.  
 XX  
 SQ Sequence 392 AA;

Query Match 42.9%; Score 860.5; DB 21; Length 392;  
 Best Local Similarity 44.2%; Pred. No. 7.1e-53;  
 Matches 156; Conservative 45; Mismatches 105; Indels 47; Gaps 6;

QY 7 AALGLPLLLLP-----PAPRA--KKPTPHRCGLVDKFNQGMVDTAKKNFPGG 56  
 Db 12 avlwglsiflnlpplwlgpspppqspppqphtctrglyvdsfnkjetrltlnf 71  
 QY 57 NTAMEKTLKSYSESEIRLEILEGLCESSDFECNOMLEAOEHLAMWLQAKSEYPDLEF 116  
 Db 72 ntaweenlskykdsedrlyevlegvcskdsdfechrllleaelveswfwfkqgpadlf 131  
 QY 117 EMFCVKTLYKCCSPGTYPDCLACQCGSQRCGSGNGHSGDGRQDSCRCRMKCYGPL 176  
 Db 132 qwlcdslkiccpagltfpcplpcpggtterpcggygcegegrtgrgshcdcaqagygaa 191

```

0Y 177 CTCDKMDYFSSLRNETHSICACADESCKTSGLNRRCGEEVGMVLDEBACADVDECAA 236
Db 192 cqqcqlgyfseearnashlvcsacfcfpacarspspeensclqckxywalhhlkcad----- 245
0Y 237 EPPCSAQFCCKMANGSVTCBECOSSCVCGCGEGPGNCKEISCYAREHGQCADVDECSL 296
Db 246 -----cakackycgmagpgtckcspgyqvgvsgkcldvdece- 282
0Y 297 AEKTCVRKENCNYMPSGVCPDGEFEETDACP--PAEA-----EATEGE 342
Db 263 -tevcpgenkbcentegaytrictdeeygkqmgievlkbpdesagffsemlde 334

```

RESULT	38
ABB65688	
ID	ABB65688 standard; Protein; 374 AA.
XX	
AC	ABB65688;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 23856.

KW Drosophila; developmental biology; cell signalling; insecticide,  
 KW pharmaceutical.  
 KW  
 OS Drosophila melanogaster.  
 XX  
 XX  
 PN W0200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EM;  
XX  
XX WPI: 2001-656860/75.  
DR N-PSDB: ABL09791.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
XX Disclosure: SEQ ID NO 23856; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent is  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention  
CC is useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB161176-AB163511), expressed DNA  
CC sequences (AB101840-AB161175) and the encoded proteins  
CC (AAB57737-AB572072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://ftp.wipo.int/pub/published_pat_sequences).

Query Match	33.4%	Score 670;	DB 22;	Length 374;
Best Local Similarity	40.5%;	Pred. No. 1.7e-39;		
Matches 122; Conservative	48;	Mismatches 109;	Indels 22;	Gaps 8

```

Oy 28 PTPCHRCRGLVDKFNOCAMVDTAKNFEGCGTAWMEKTLKYESSEIRLLEILEGCESSD 87
    ||| ||| : : : ||:||||: | :||:|:|:| ||| :
Db 28 pppcraactqlvssfragl-ertkghnagdcraweekrlsrkxhsevrllvelqeklcsge 86

```

```

0y      88 F--ECNOMLEAEDEHEAAMWLDLKSSEYDLPFMFCVNTLKVOCSGPTGPPOLACGG 143
        | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      87 vinkdchmininehealleedfifkktgespolgswlctdqtlvccpnlytpdclic--- 143

0y      144 SORPSCNGHCSGDGRSGDCSRCHMGYOGLCTDMDGYFSSLRNETHSICTADDESC 203
        | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      144 --tegnqckckgdrtfkngnkcakcdpyagpmnecegpelhesfrdekhlldtgcnaac 201

0y      204 KT--CSGLTINDCGCEGWULD-EGACVDVDECAAE--PPCSAAOFCKNANGSYTCEE 258
        | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      202 geggactggypksctrckckgywmsdeagcvdlneclbeqrlnpcrppqfcvnmegsfscle 261

0y      259 CDSGCVGTGEGBPGNCKECISGVAREHGQCADDECSLAEKTCVRKNENCYNTPGSTVCV 318
        | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      262 cdrrscdgcdgagpdmcrckaaygelkegkchdis----aeq-----rsnyvsftmltyfgm 314

0y      319 C 319
        |
Db      315 c 315

```

RESULT	39
AAV76114	
ID	AAV76114 standard; Protein; 242 AA
XX	
AC	AAV76114;
XX	
DT	27-MAR-2000 (first entry)

DE	Rat HT glycoprotein homologue, SEQ ID NO:353.
XX	
KW	Skin; dermal papillae; keratinocyte; neonatal foreskin fibroblast;
KW	embryonic skin cells; keratinocyte stem cell; transit amplifying cell;
KW	secreted; transmembrane; inflammation; cancer; neurological disease;
KW	angiogenesis; tumour vascularisation; growth disorder;
KW	developmental disorder; skin wound; hair follicle disorder;
KW	anti-inflammatory; cytostatic; neuroprotective; vulvury.

OS	Rattus sp.	
XX		
PN	W09955865-A1.	
PD	04-NOV-1999.	
XX		
PF	29-APR-1999;	99WO-NZ00051.
XX		
PR	29-APR-1999;	98US-0069726
ER	09-NOV-1999;	98US-0188930
XX		

PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX  
PI Strachan L, Sleeman M, Watson JD, Onrust R, Kumble A, Murison JG.  
XX  
WPI: 2000-072177/06.  
XX  
DR N-PSDB; AA261822.  
DR

PR Novel polynucleotides useful for the treatment of various conditions  
PT including wounds and cancer -  
XX  
PS Claim 4; Page 222; 235pp; English.

CC The invention relates to novel nucleic acid sequences derived from rat  
CC dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,  
CC and mouse embryonic skin, keratinocyte stem cells and transit amplifying  
CC cells. Polypeptides of the invention may be used to treat inflammation,  
CC cancer and neurological diseases. The proteins may be used to stimulate  
CC the growth and motility of keratinocytes, to inhibit the growth of  
CC cancer cells, to modulate angiogenesis and tumour vascularisation, to  
CC modulate skin inflammation, to modulate epithelial cell growth and to  
CC inhibit binding of HIV-1 to leukocytes. The invention may also be used  
CC to treat growth and developmental defects, skin wounds and hair follicle  
CC disorders. Sequences AAV75942-V76123 represent polypeptides encoded



PT Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 PS Claim 20: Page 241: 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoietic regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 CC  
 XX  
 XX Sequence 140 AA:  
 SQ  
 Query Match 20.2%; Score 406; DB 22; Length 140;  
 Best Local Similarity 100.0%; Fred. No. 2.3e-21;  
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 258 ECDSCVCGTGEAGNCKECISGAREHGACADVDSCSLAEKTCVRKNENCYWTPGSSYC 317  
 DB 7 ecdscevgctgeagncckecisgarehgacadvdcslaecktcrknencynwtpgssyc 66  
 QY 318 VCPDGFEEET 326  
 DB 67 vcpdgfeet 75  
 RESULT 42  
 AAR05222  
 ID AAR05222 standard; protein; 2189 AA.  
 XX  
 XX AAR05222;  
 AC  
 XX  
 XX 02-AUG-1990 (first entry)  
 DE Antigen GX5401FL encoded by *Elmeria tenella* genomic DNA.  
 XX  
 XX *Elmeria tenella*; antigen GX5401FL; antigen GX5401; avian coccidiosis.  
 XX  
 OS *Elmeria tenella*.

XX  
 PN WO9000403-A.  
 XX  
 PD 25-JAN-1990.  
 XX  
 PF 05-JUL-1989; 89WO-US02918.  
 XX  
 PR 05-JUL-1988; 88US-0215162.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 PI Anderson DM, Mcandliss RJ, Strausberg SL, Strausberg RL;  
 DR WPI: 1990-051586/07.  
 DR N-PSDB: AA003324.  
 XX  
 PT Cloned gene or fragment encoding antigenic protein -  
 PT which binds with antibodies against avian coccidia, and  
 PT transformed cells used in vaccine  
 XX  
 PS Claim 10; Page 93; Fig 14; 134pp; English.  
 XX  
 CC It is encoded by an open reading frame contained within the sequence of  
 CC clone 533 which was derived from an *E. tenella* genomic library screened  
 CC with radioactively labelled cDNA encoding the GX5401 antigen. It is of  
 CC about 250 Kd. It carries several repeated peptide sequences and  
 CC is rich in cysteine residues. The open reading frame also encodes a  
 CC potential signal sequence for protein secretion. Also new are an  
 CC expression vector contg. cloned gene, and host cells transformed with  
 CC the vector. The transformed cells are used in a vaccine to immunise  
 CC birds against avian coccidiosis. By labelling the peptides, they can be  
 CC used as a type-specific probe. May also be used in an assay to detect  
 CC Ab against the coccidia. The Abs are used to identify transformed cells  
 CC contg. the DNA.  
 CC  
 XX  
 XX Sequence 2189 AA:  
 SQ  
 Query Match 13.7%; Score 274.5; DB 11; Length 2189;  
 Best Local Similarity 27.5%; Fred. No. 7.7e-11;  
 Matches 81; Conservative 28; Mismatches 81; Indels 105; Gaps 15;  
 QY 126 VCCSPGTYGPBDCIACGG-----SQPCSGNGHSGSGSRGDSGSCRC 168  
 DB 318 vctn--tegsytackegfsgfegaagcadvdecanspdaahasca--ntegsyvctc 372  
 QY 169 HMGYGPLECTDCMDYGFSSLRNETHSICACDESKTCGSLNRD-----CGRECVGWTD 224  
 DB 373 npyeap-----ssdgnackdvde---caagtaechvsqcvnvdgsyec-hclegfld 423  
 QY 225 EGACVVDDECAEAEPPECSAOFCKNANGSYTCE-----EC----- 259  
 DB 424 gkxcsdvdecaaeaapcgnthlnltlgsyececkdgyhmgngasddideceasteip 483  
 QY 260 -DSSCVGCTGE-----GPG-----NCKE-----CIGS 280  
 DB 484 encncvntegsfsleakpyelvdgcvkldfcaargacnslahckenpegtaicctclag 543  
 QY 281 YARE--HGOCADVDECSLAERTVKRKNEN--CYNPGSYVCGCPDGFETEDAC 330  
 DB 544 ysgdgtaghdiddidcc-laendctpadggjlecnlvgsytckcaagyqdgdnsc 597  
 RESULT 43  
 AAY32346  
 ID AAY32346 standard; Protein; 637 AA.  
 XX  
 XX AAY32346;  
 AC  
 XX  
 XX 28-FEB-2000 (first entry)  
 DE Mouse cell surface receptor C1qRP.  
 XX  
 XX

KW C1qR; receptor; cell surface; transmembrane; glycoprotein; mouse;  
 KW host defence; infection; HIV; immunodeficiency; therapy;  
 KW immunostimulant; phagocytosis; signal transduction.  
 XX  
 OS Mus musculus.  
 XX  
 PN MO9955839-A1.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PF 29-APR-1999; 99WO-US09335.  
 XX  
 PR 30-APR-1998; 98US-0071386.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Tenner AJ, Nepomuceno RR;  
 DR WPI: 2000-062021/05.  
 DR N-PSDB; AA234990.  
 XX  
 PT A new cell surface receptor protein used as a prophylactic for  
 PT individuals at risk from infection, e.g. HIV -  
 PS Claim 2; Fig 8; 49pp; English.  
 XX  
 XX This sequence represents a novel mouse cell surface transmembrane  
 CC glycoprotein receptor, designated C1qR, as deduced from DNA (see  
 CC AA234990) isolated from a genomic library. C1qR plays a role in  
 CC stimulating the classic complement component of the immune system,  
 CC specifically in stimulating phagocytosis in cells without a  
 CC concomitant increase in inflammation. Methods for detecting novel  
 CC ligands for C1qR, including those which function as agonists or  
 CC antagonists, are provided, as well as methods of determining  
 CC compositions, which effect the formation of an affinity complex  
 CC between the C1qR and its ligand, and for determining compositions  
 CC which modulate signal transduction via the C1qR. Transgenic  
 CC animals can be created to aid in the study of the role of C1qR  
 CC during growth and metabolism and as a model for disease states in  
 CC which the normal level of C1qR is effected. The ability to  
 CC regulate the phagocytic capacity of myeloid cells via the  
 CC regulation of cell surface expression and function of C1qR will  
 CC be valuable as a prophylactic treatment of individuals at risk  
 CC from infection, particularly those with genetic immunodeficiencies,  
 CC HIV infection, or undergoing cancer chemotherapy or high risk  
 CC surgery.  
 CC  
 XX  
 XX Sequence 637 AA;  
 SQ

Query Match 13.3%; Score 266.5; DB 21; Length 637;  
 Best Local Similarity 27.0%; Pred. No. 7.6e-11;  
 Matches 111; Conservative 40; Mismatches 119; Indels 141; Gaps 28;

QY 12 LPLLLPPAPAEARK--PRPC-----HRCRGLVDFN-OGWDTAKKNGCGNTAME 61  
 DB 143 LLLLSLtlphshlpkwnspcpgpeapnsliegflckfnfkmcprla1g-gpvrlyt 201  
 QY 62 ---EKTLASKYSSSEIRLEILEGLC-----ESSDFECNMLEADOEHLLEAMWLOLKSEY 112  
 DB 202 tpfatstsslea--vpfssvanvacgdeaksethlfline-----kt 241  
 QY 113 PDLEWPCVTKLWCKSPGTGPPCLACGGSGRPS--GNGHCGSDGSRQDGS--CQCH 169  
 DB 242 pglthw-----gsagplcvspkfg-----cstfnngcqqgdcfegdgstfrcgr 285  
 QY 170 MGQY-----GPLCTD-----CMDGYFSSLRNETHSI--- 195  
 DB 286 pglfllldlvcaarnpsncpctggumchavplsenytcrcpsgyqd--sqvncvold 344  
 QY 196 -C--TACDESKTCSGLTNRDCGCEVGVWLD---EGACVYDDECAAPPCSAQFCKN 249  
 DB 345 ecqdspecdcvntly--stfnc-ecwvygqpsgpkceacedvdecaaanpc--aqgcin 399

QY 250 ANGSYCEECDCSSGVCTGEGPGNCKECISGV---AREHGCAVDDECSLAE-KTCVRKN 305  
 DB 400 tdgsfyc-----ske---glyvsgegstctgedtdesdrnrc-----d 437  
 QY 306 ENCYNTPGSSVYCVCPDGFETEDACVPPAEAEATEG---ESP70LPSPRED 352  
 DB 438 slcftdgsfrcgpcppwye-----lapngvfcsgrgtvfse1parppked 482

RESULT 44  
 AAY79186  
 ID AAY79186 standard; Protein: 644 AA.  
 XX  
 AC AAY79186;  
 XX  
 DT 19-JUN-2000 (first entry)  
 XX  
 DE Haematopoietic stem cell specific protein.  
 XX  
 KW Haematopoietic stem cell; immune system disorder;  
 KW Leukaemia; antileukaemic; immunomodulator; therapy; mouse.  
 XX  
 XX Mus musculus.  
 OS  
 PN MO200011168-A2.  
 XX  
 PD 02-MAR-2000.  
 XX  
 PF 20-AUG-1999; 99WO-US19052.  
 XX  
 PR 21-AUG-1998; 98US-0138132.  
 XX  
 PA (UYPR-) UNIV PRINCETON.  
 XX  
 PI Lemischka I, Moore K;  
 DR WPI: 2000-237650/20.  
 DR N-PSDB; AA294127.  
 XX  
 PT Hematopoietic stem cell signaling proteins modulating replication and  
 PT differentiation for treating immune system disorders and leukaemia -  
 PS Claim 21; Page 236-238; 256pp; English.  
 XX  
 XX The present sequence is that of a mouse haematopoietic stem cell  
 CC (HSC) specific protein. It is an example of claimed HSC-specific  
 CC proteins (see AAY79176-93) predicted from novel isolated HSC-specific  
 CC nucleic acids (see AA294077-131). The HSCs are especially primitive  
 CC HSCs (PHSCs) such as umbilical cord cells, bone marrow cells and  
 CC foetal liver cells. The encoded proteins are growth factors, and  
 CC transcription factors, splicing factors, capping factors, transport  
 CC proteins, translation factors or replication factors that modulate  
 CC HSC activity, especially differentiation or replication. The  
 CC invention provides claimed methods: for identifying PHSC-specific  
 CC nucleic acids: for generating a stem cell/progenitor cell from  
 CC PHSCs: for identifying the presence of a PHSC in a sample; for  
 CC identifying the presence in a sample of a compound that modulates  
 CC HSC activity; for using such a compound to treat an immune system  
 CC condition, especially leukaemia; for introducing exogenous nucleic  
 CC acid into a HSC; and for ex vivo expansion of HSCs. Also claimed  
 CC are vectors, host cells, and an antibody that specifically binds a  
 CC an HSC-specific protein.  
 CC  
 XX  
 XX Sequence 644 AA;  
 SQ

Query Match 13.3%; Score 266.5; DB 21; Length 644;  
 Best Local Similarity 27.0%; Pred. No. 7.7e-11;  
 Matches 111; Conservative 40; Mismatches 119; Indels 141; Gaps 28;

QY 12 LPLLLPPAPAEARK--PRPC-----HRCRGLVDFN-OGWDTAKKNGCGNTAME 61  
 DB 143 LLLLSLtlphshlpkwnspcpgpeapnsliegflckfnfkmcprla1g-gpvrlyt 201  
 QY 62 ---EKTLASKYSSSEIRLEILEGLC-----ESSDFECNMLEADOEHLLEAMWLOLKSEY 112  
 DB 202 tpfatstsslea--vpfssvanvacgdeaksethlfline-----kt 241  
 QY 113 PDLEWPCVTKLWCKSPGTGPPCLACGGSGRPS--GNGHCGSDGSRQDGS--CQCH 169  
 DB 242 pglthw-----gsagplcvspkfg-----cstfnngcqqgdcfegdgstfrcgr 285  
 QY 170 MGQY-----GPLCTD-----CMDGYFSSLRNETHSI--- 195  
 DB 286 pglfllldlvcaarnpsncpctggumchavplsenytcrcpsgyqd--sqvncvold 344  
 QY 196 -C--TACDESKTCSGLTNRDCGCEVGVWLD---EGACVYDDECAAPPCSAQFCKN 249  
 DB 345 ecqdspecdcvntly--stfnc-ecwvygqpsgpkceacedvdecaaanpc--aqgcin 399

Db 143 lldlsitphshlpkwhespcgtpeapgnslgflcklnfkymcrplalg-gpgrvtyt 201  
 QY 62 ---EKTLSKYESSEIRLLEILEGLC-----ESSDFECNOMLEAQBHLPAWMLQLKSEY 112  
 Db 202 tpfqatstsllea--vpfasvanvacgdeaksethlylcn-----kt 241  
 QY 113 PDLEFMPCKYKTLKVCSPGTYGPDCLACGGSGRPCS-GNGHSGDGSROGDGS--CRCH 169  
 Db 242 pglfhw-----gssgplcvspkfg-----csfnngcgqdcfeggdgsfrgcgr 285  
 QY 170 MGYQ-----GPLCTD-----CMDGYFSSLRNETHSI--- 195  
 Db 286 pglfllldlvtcasrnpssnptctggmchsvplsenytrcpsyqld-sgvyhcvld 344  
 QY 196 -C--TACDESKCTKSGLTNRDCECEVGNWLD--EGACVDVDECAEPSPCSAOFCKN 249  
 Db 345 ecqdspcagdcvntlg--sfhc-ecwvygypspkceacedvdecaanspc--agqcin 399  
 QY 250 ANGSTYCEECDDSSCVGCTGEGPNCCKECLISGY--AREHGQCADVDECSLAE-KTCVRKN 305  
 Db 400 tdgsfyc-----scke-----gylvsgedstqcedidcscdargnpc---d 437  
 QY 306 ENCYNTPGSGVVCVCPDGFETEDACVPAEAETEG---ESPQLPSRED 352  
 Db 438 slcfnldgsfrgcgppgwe-----lapngvfcsrtglvtselparpqked 482

## RESULT 45

AA79193 standard; protein; 644 AA.

AA79193;

19-JUN-2000 (first entry)

DE Haematopoietic stem cell specific protein AA4.  
 AA4: haematopoietic stem cell; immune system disorder;  
 KW leukaemia; antileukaemic; immunomodulator; therapy; mouse.  
 OS Mus musculus.  
 FH Key Location/Qualifiers  
 FT Peptide 1..19 /note= "signal peptide"  
 FT Protein 20..644 /note= "mature protein"  
 FT Domain 35..129 /note= "C-type lectin carbohydrate recognition d"  
 FT Domain 575..597 /note= "transmembrane domain"  
 FT Region 282..297 /note= "EGF-like repeat"  
 FT Region 325..340 /note= "EGF-like repeat"  
 FT Region 451..464 /note= "EGF-like repeat"  
 FT Region 343..364 /note= "calcium-binding EGF motif"  
 FT Region 382..406 /note= "calcium-binding EGF motif"  
 FT Region 424..449 /note= "calcium-binding EGF motif"  
 FT Region /note= "calcium-binding EGF motif"

MO200011168-A2.

02-MAR-2000.

20-AUG-1999; 99MO-US19052.

21-AUG-1998; 98US-0138132.

(UYPR-) UNIV PRINCETON.

XX Lemischka I, Moore K;  
 PI  
 DR WPI: 2000-237650/20.  
 DR N-PSDB; AA294131.  
 XX  
 PT Hematopoietic stem cell signaling proteins modulating replication and  
 PT differentiation for treating immune system disorders and leukaemia -  
 PS Claim 21; Fig 9; 256pp; English.

CC The present sequence is that of mouse haematopoietic stem cell  
 CC (HSC) specific protein AA4, a type I transmembrane protein that  
 CC shows homology to C1qR, the human receptor for complement C1q.  
 CC AA4 is a preferred example of claimed HSC-specific proteins (see  
 CC AA79176-93) predicted from novel isolated HSC-specific nucleic acids  
 CC (see AA294077-131). The HSCs are especially primitive HSCs (PHSCs)  
 CC such as umbilical cord cells, bone marrow cells and foetal liver  
 CC cells. The encoded proteins are growth factors, transcription  
 CC factors, splicing factors, capping factors, transport proteins,  
 CC translation factors or replication factors that modulate HSC  
 CC activity, especially differentiation or replication. The invention  
 CC provides claimed methods: for identifying PHSC-specific nucleic  
 CC acids; for generating a stem cell/progenitor cell from PHSCs; for  
 CC identifying the presence of a PHSC in a sample; for identifying the  
 CC presence in a sample of a compound that modulates HSC activity; for  
 CC using such a compound to treat an immune system condition,  
 CC especially leukaemia; for introducing exogenous nucleic acid into a  
 CC HSC; and for ex vivo expansion of HSCs. Also claimed are vectors,  
 CC host cells, and an antibody that specifically binds an HSC-specific  
 CC protein.

SQ Sequence 644 AA;

Query Match 13.3%; Score 266.5; DB 21; Length 644;  
 Best Local Similarity 27.0%; Pred. No. 7.7e-11;  
 Matches 111; Conservative 40; Mismatches 119; Indels 141; Gaps 28;

QY 12 LPDLLLPAPAEARK--PTPC-----HRCGLVDKFN-QGMVDTAKNFGGNTAME 61  
 Db 143 lldlsitphshlpkwhespcgtpeapgnslgflcklnfkymcrplalg-gpgrvtyt 201  
 QY 62 ---EKTLSKYESSEIRLLEILEGLC-----ESSDFECNOMLEAQBHLPAWMLQLKSEY 112  
 Db 202 tpfqatstsllea--vpfasvanvacgdeaksethlylcn-----kt 241  
 QY 113 PDLEFMPCKYKTLKVCSPGTYGPDCLACGGSGRPCS-GNGHSGDGSROGDGS--CRCH 169  
 Db 242 pglfhw-----gssgplcvspkfg-----csfnngcgqdcfeggdgsfrgcgr 285  
 QY 170 MGYQ-----GPLCTD-----CMDGYFSSLRNETHSI--- 195  
 Db 286 pglfllldlvtcasrnpssnptctggmchsvplsenytrcpsyqld-sgvyhcvld 344  
 QY 196 -C--TACDESKCTKSGLTNRDCECEVGNWLD--EGACVDVDECAEPSPCSAOFCKN 249  
 Db 345 ecqdspcagdcvntlg--sfhc-ecwvygypspkceacedvdecaanspc--agqcin 399  
 QY 250 ANGSTYCEECDDSSCVGCTGEGPNCCKECLISGY--AREHGQCADVDECSLAE-KTCVRKN 305  
 Db 400 tdgsfyc-----scke-----gylvsgedstqcedidcscdargnpc---d 437  
 QY 306 ENCYNTPGSGVVCVCPDGFETEDACVPAEAETEG---ESPQLPSRED 352  
 Db 438 slcfnldgsfrgcgppgwe-----lapngvfcsrtglvtselparpqked 482

## RESULT 46

AA78887 standard; protein; 1121 AA.

AA78887;



XX 14-DEC-2001 (first entry)  
XX Human fibrillin 3.  
DE  
XX  
XX Human; fibrillin; gene therapy; Marfan's syndrome.  
OS Homo sapiens.  
XX JP2001245664-A.  
XX  
XX 11-SEP-2001.  
XX  
XX 06-MAR-2000; 2000JP-0060009.  
XX  
XX 06-MAR-2000; 2000JP-0060009.  
XX  
XX (KAZU-) 2H KAZUSA DNA KENKYUSHO.  
XX  
XX WPI: 2001-610073/70.  
XX N-PSDB; AA169907.  
XX  
XX  
XX New human fibrillin 3 gene for treating and preventing diseases showing  
PT symptoms similar to Marfan's syndrome -  
XX  
XX  
XX Claim 1: Page 11-14; 17pp: Japanese.  
XX  
XX The present sequence is the protein sequence for human fibrillin 3.  
CC fibrillin 3 and its coding sequence can be used for the treatment and the  
CC prevention of diseases showing symptoms similar to Marfan's syndrome.  
XX  
XX Sequence 1121 AA:

Query Match 13.0%; Score 260; DB 22; Length 1121;  
Best Local Similarity 23.2%; Pred. No. 4e-10;  
Matches 96; Conservative 32; Mismatches 120; Indels 166; Gaps 20;

QY 69 ESSEIR-----LLEFLLEGICSSDPEFCNQMLEAQEHEHLEAMWLQKSEYPLLEFPCVKTL 124  
DB 603 echdlitgqpcfaevlqmcrrs-----lssse-----avtra 634  
QY 125 KVCSPGT-YGPDLAC-----QGGSQRPCSGNGHSGDQ--SR 160  
DB 635 ecccggyrgwgcclcpqlgtsayrkicphgsgytaegrdvdecmhlalhahgectins 694  
QY 161 QGSGSCRHNGY-QGPLCTDCMCGYFSSLRNETHSITCTACDESKT-----CS----- 207  
DB 695 lgsfrchcgggyrpdattatctid-----ndecsqvbkpctficknktgssflscprgyl 748  
QY 208 ----GLTNRCGE-----CEVGVNLDGACVDVDECAAEPPPC 241  
DB 749 leedgtrckkldectrqhncqflcvntvgaftccrppgftqhqaefndecsaagpqc 808  
QY 242 SAAQFCKNANGSYCEEC-----SSCVGCTG---EGFGNCK-----ECIS 279  
DB 809 gahghchmnpqstfrc-echqgftlvssghgcedvnedghrhcqhngqnlgyrscpcp 867  
QY 280 GYAR--EHGQCAVDVDECSLAETKCVKRNENCYTNGSYVCVCPDGF-----ETEDA 329  
DB 868 gftcghsqwagcvdeneacalspctc--gsaascrltlgfrfcvcpsgfaldqagcqvde 925  
QY 330 CV-----PPAEAEATGEG-----SPQLPSPREDL 353  
DB 926 cagrrpcpsyscantpvgflicgpcpgyfragqghcvsqglfispqpdtpkde 979

RESULT 47  
AAAY70551  
ID AAY70551 standard; Protein; 1208 AA.  
XX  
XX AC AAY70551;  
XX

DT 04-JUL-2000 (first entry)  
XX  
XX Human latent transforming growth factor-beta binding protein 3 (1).  
DE  
XX  
XX Human latent transforming growth factor-beta binding protein 3; hlrp-3;  
KW TGF-beta inhibitor; proliferative; anti-proliferative; cytosolic;  
KW cardiant; anti-inflammatory; cerebroprotective; immunosuppressive;  
KW thrombolytic; osteopathic; vulnerrary; tranquilizer; antibacterial; PAI-1  
KW plasminogen activator inhibitor-1; thrombomodulin; tissue growth; tumour  
KW prevention; treatment; cancer; fibrosis; dilated cardiomyopathy; injury;  
KW osteoporosis; myocardial infarction; congestive heart failure; sepsis;  
KW thrombosis; stroke; systemic inflammatory response syndrome;  
KW septic shock; sepsis syndrome; multiple organ dysfunction syndrome;  
KW atherosclerotic plaque rupture.  
XX  
XX Homo sapiens.  
XX  
XX WO200012551-A1.  
XX  
XX 09-MAR-2000.  
XX  
XX 30-AUG-1999; 99WO-US19436.  
XX  
XX 01-SEP-1998; 98US-0098766.  
XX  
XX (ELIL ) LILLY & CO ELI.  
XX  
XX Edmonds BT;  
XX  
XX WPI: 2000-256589/22.  
XX N-PSDB; AA251928.  
XX  
XX  
XX Human latent transforming growth factor (TGF)-beta binding protein 3,  
PT nucleic acids and vectors useful as modulators of TGF-beta, for  
PT inhibiting tissue or tumor growth, and treating e.g. osteoporosis and  
PT myocardial infarction -  
XX  
XX  
XX Claim 1: Page 61-66; 78pp: English.

XX The present sequence is a human latent transforming growth factor  
CC (TGF)-beta binding protein 3 (hlrbp-3) which is expressed in various  
CC tissues e.g. cartilage, bone, brain, kidney, ovary, pancreas, stomach  
CC and spleen. The protein is often co-expressed with TGF-beta and functions  
CC as its inhibitor. hlrp-3 is useful for inhibiting or stimulating tissue  
CC growth in vitro or in vivo and for inhibiting tumour growth. The hlrp-3;  
CC or its homologues, and antisense nucleic acid sequences can be used to  
CC regulate TGF-beta activity, especially plasminogen activator inhibitor-1  
CC expression, activity or secretion, thrombomodulin expression or  
CC activity, TGF-beta secretion and cellular proliferation. Modulation of  
CC hlrp-3 is useful for prevention and/or treatment of diseases arising  
CC from cellular effects induced by TGF-beta, especially cancer, fibrosis,  
CC osteoporosis, myocardial infarction, congestive heart failure, dilated  
CC cardiomyopathy, deep venous thrombosis, disseminated intravascular  
CC thrombosis, stroke, sepsis, injuries involving major tissue damage and  
CC trauma, systemic inflammatory response syndrome, sepsis syndrome,  
CC septic shock, multiple organ dysfunction syndrome, and atherosclerotic  
CC plaque rupture.  
XX  
XX Sequence 1208 AA:

Query Match 12.8%; Score 256.5; DB 21; Length 1208;  
Best Local Similarity 26.8%; Pred. No. 7.6e-10;  
Matches 84; Conservative 32; Mismatches 91; Indels 107; Gaps 20;

QY 128 CSPG-----TYGPPDLACQGSQRPCS-GNGHCSGDSRGDSCRHNGYQ----- 173  
DB 553 cnpqyrszhpqhyrcvdenecae---epcgprgiclmty---gsynhcmrgrylhvgag 606  
QY 174 GPLCTD-----CMDGIFS-----SLRNETHSITCTACDESKTCSGL 209  
DB 607 grscvdlneacaphlcgdggfclnfpghkcnycpysyrlkasrpvcddide-crdbssc 665

```

QY 210 TNRDGE-----CEWGW-VLDEGACVDDECAAEPPPCSAOFCNKANGSYTC--- 256
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 666 pdgkceknkpgsfkclacqgyrsqggagacdvneac-aeqspcspg-wcenlpgsftrctca 723
QY 257 -----EECDSSCV---GCTGEGSPGNCK-ECISGY--AREHGOCADVDECS 295
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 724 qgyapapdgfrcscldvdeceagdvcdnglcsnlpqsfqcgclsgylsrdrshcdedec 783
QY 296 LAEKTCVRKNEKYNTPGSGYVCVCPDGF-----EETEDACVPAEAE---- 337
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 784 f-paaci--ggdcintngsryclcpqghrlvggrkcgdldecsgdpslclphgacknlg 840
QY 338 -----ATEGESPTQ 346
      | | | | |
Db 841 syvcvdegfptq 854

RESULT 48
AA70554
ID AAY70554 standard; Protein; 1257 AA.
XX
AC AAY70554;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human latent transforming growth factor-beta binding protein 3 (III).
XX
KW Human latent transforming growth factor-beta binding protein 3; hLTBP-3;
KW TGF-beta inhibitor; proliferative; anti-proliferative; cyostatic;
KW cardiant; anti-inflammatory; cerebroprotective; immunosuppressive;
KW thrombolytic; osteopathic; vulnerrary; tranquillizer; antibacterial; PAI-1;
KW plasminogen activator inhibitor-1; thrombomodulin; tissue growth; tumour;
KW prevention; treatment; cancer; fibrosis; dilated cardiomyopathy; injury;
KW osteoporosis; myocardial infarction; congestive heart failure; sepsis;
KW thrombosis; stroke; systemic inflammatory response syndrome;
KW septic shock; sepsis; systemic inflammatory response syndrome;
KW atherosclerotic plaque rupture.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 189
FT /label= Unknown
XX
PN WO200012551-A1.
XX
PD 09-MAR-2000.
XX
PE 30-AUG-1999; 99WO-US19436.
XX
PR 01-SEP-1998; 98US-0098766.
XX
PA (ELIL ) LILLY & CO ELI.
XX
PI Edmonds BT;
XX
DR WPI: 2000-256589/22.
XX
PT Human latent transforming growth factor (TGF)-beta binding protein 3,
PT nucleic acids and vectors useful as modulators of TGF-beta, for
PT inhibiting tissue or tumor growth, and treating e.g. osteoporosis and
PT myocardial infarction -
XX
PS Claim 1, Page 69-74; 78pp; English.
XX
CC The present sequence is a human latent transforming growth factor
CC (TGF)-beta binding protein 3 (hLTBP-3) which is expressed in various
CC tissues e.g. cartilage, bone, brain, kidney, ovary, pancreas, stomach
CC and spleen. The protein is often co-expressed with TGF-beta and functions
CC as its inhibitor. hLTBP-3 is useful for inhibiting or stimulating tissue
CC growth in vitro or in vivo and for inhibiting tumour growth. The hLTBP-3,
CC or its homologues, and antisense nucleic acid sequences can be used to
CC regulate tgf-beta activity, especially plasminogen activator inhibitor-1

```

```

CC expression, activity or secretion, thrombomodulin expression or
CC activity, TGF-beta secretion and cellular proliferation. Modulation of
CC hLTBP-3 is useful for prevention and/or treatment of diseases arising
CC from cellular effects induced by TGF-beta, especially cancer, fibrosis,
CC osteoporosis, myocardial infarction, congestive heart failure, dilated
CC cardiomyopathy, deep venous thrombosis, disseminated intravascular
CC thrombosis, stroke, sepsis, injuries involving major tissue damage and
CC trauma, systemic inflammatory response syndrome, sepsis syndrome,
CC septic shock, multiple organ dysfunction syndrome, and atherosclerotic
CC plaque rupture.
XX
SQ Sequence 1257 AA:

Query Match 12.8%; Score 256.5; DB 21; Length 1257;
Best Local Similarity 26.8%; Pred. No. 7.9e-10;
Matches 84; Conservativity 32; Mismatches 91; Indels 107; Gaps 20:

QY 128 CSPG-----TYGPDCLACGGGSGRPCS-GNGHSGSGSRGGDSRCRHMGYO----- 173
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 602 cnpyrshpqhrycvdvnecea---epcgyrglcmntg---gsynchnrgyrlhvqag 655
QY 174 GPLCTD-----CMDCYFS-----SIRNETHSICACDSECKTCGL 209
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 656 grscvdlnecaakphlccgdgfcinfphkcnypgyrlkaerpyvcedide-ctdpssc 714
QY 210 TNRDGE-----CEWGW-VLDEGACVDDECAAEPPPCSAOFCNKANGSYTC--- 256
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 715 pdgkceknkpgsfkclacqgyrsqggagacdvneac-aeqspcspg-wcenlpgsftrctca 772
QY 257 -----EECDSSCV---GCTGEGSPGNCK-ECISGY--AREHGOCADVDECS 295
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 773 qgyapapdgfrcscldvdeceagdvcdnglcsnlpqsfqcgclsgylsrdrshcdedec 832
QY 296 LAEKTCVRKNEKYNTPGSGYVCVCPDGF-----EETEDACVPAEAE---- 337
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 833 f-paaci--ggdcintngsryclcpqghrlvggrkcgdldecsgdpslclphgacknlg 889
QY 338 -----ATEGESPTQ 346
      | | | | |
Db 890 syvcvdegfptq 903

RESULT 49
AAB61460
ID AAB61460 standard; protein; 1260 AA.
XX
AC AAB61460;
XX
DT 04-APR-2001 (first entry)
XX
DE TANGO 275 mature protein.
XX
KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;
KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;
KW pancreatic; skeletal; muscle.
XX
OS Homo sapiens.
XX
PN WO200100672-A1.
XX
PD 04-JAN-2001.
XX
PE 29-JUN-2000; 2000WO-US18184.
XX
PR 29-JUN-1999; 99US-0342687.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;
XX
DR WPI: 2001-050127/06.
XX

```

PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g., idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g., jaundice) =

PS Disclosure; Page 236-239; 262pp; English.

CC The present invention relates to cDNAs encoding TANGO 244,  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC The amino acid sequences of TANGO 244, TANGO 246, TANGO 275,  
CC TANGO 300 and MANGO 245 are set forth in SEQ. ID. NO. 1, 2, 3, 4,  
CC 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23,  
CC 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40,  
CC 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57,  
CC 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74,  
CC 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,  
CC 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106,  
CC 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119,  
CC 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132,  
CC 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145,  
CC 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158,  
CC 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171,  
CC 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184,  
CC 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197,  
CC 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210,  
CC 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223,  
CC 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236,  
CC 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249,  
CC 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262,  
CC 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275,  
CC 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288,  
CC 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301,  
CC 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314,  
CC 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327,  
CC 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340,  
CC 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353,  
CC 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366,  
CC 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379,  
CC 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392,  
CC 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405,  
CC 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418,  
CC 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431,  
CC 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444,  
CC 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457,  
CC 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470,  
CC 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483,  
CC 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496,  
CC 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509,  
CC 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522,  
CC 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535,  
CC 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548,  
CC 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561,  
CC 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574,  
CC 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587,  
CC 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600,  
CC 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613,  
CC 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626,  
CC 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639,  
CC 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652,  
CC 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665,  
CC 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678,  
CC 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691,  
CC 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704,  
CC 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717,  
CC 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730,  
CC 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743,  
CC 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756,  
CC 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769,  
CC 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782,  
CC 783, 784,

CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.

**SQ** Sequence 1260 AA;

Query Match	12.8%	Score 256.5	DB 22	Length 1260
Best Local Similarity	26.8%	Pred. No. 8e-10	91	Indels 107
Matches 84	Conservative 32	Mismatches		Gaps 20

```

Oy 128 CSPG-----TYPPDLACGGSGRPCS--GNHGSCSDGSRQDSCRMHGQ-----173
Db 558 cnpyrsrhqhrlycvdrneeca---epcggrgrjclcmtyt---gsynchcmryrlhvnag 611
Oy 174 GPLLTD-----CMDGRFS-----SLRNETHSITCTACDESKCTCSGL 209
Db 612 grscvcdlnecakphlccdgagfclnfpbhykcnocypryrlasrppvcedide--cridpsnc 670
Oy 210 TNRCCGE-----CEVGV-WVLEDEGACVDVDECAEPPECSAOBFCKNANSSYC---256
Db 671 pdgycenkpkpsfkclacqpyrsgsggaactdvneec--aegspcspg--wclnlpssftcta 728
Oy 257 -----EECDSSCV-----GCTGBGPNCK--ECISGY--AREHGCCADVDECS 295
Db 729 qgyvapdgrscldvdceaeagdvcdnglcsanltpbsfcgcqlsglshlstrdsheedidec 788
Oy 296 LAETKVRKNKNCNYNTGSGYCVCPDGF-----EBTEDACYPAAEA---337
Db 789 f-paactl--ggdcintnrgsyrclecpghrlvgarkcqdldcecsqdpbslclphacknlqg 845
Oy 338 -----ATGEGSPQ 346
Db 846 syvcvcdeagfptlg 859

```

RESULT	50
AAB61419	
ID	AAB61419 standard; protein; 1289 AA

AC AAB61419;

DT 04-APR-2001 (first entry)

DE Human TANGO 275 protein.

KM TANGO; MANCO; colon; inflammation; tumor; renal; liver; lung;  
KM autoimmune; allergy; cardiovascular; brain; degenerative; placental;  
KM pancreatic; skeletal; muscle.

OS Homo sapiens.

PN WO200100672-A1.

PD 04-JAN-2001.

PF 29-JUN-2000; 2000WO-US18184.

PR 29-JUN-1999; 99US-0342687.

PA (MILL-) MILLENNIUM PHARM INC.

XX Holtzman DA, Barnes TM, Fraser CC, Sharp JD,  
PI WPI: 2001-050127/06.  
XX  
DR

PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and  
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases  
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver  
PT disorders (e.g. jaundice) -

PS Claim 1; Fig 9; 262pp; English

CC The present invention relates to cDNAs encoding TANGO 244  
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.  
CC

CC The nucleic acids, proteins and protein modulators are useful for  
CC treating colonic disorders, inflammatory diseases, tumors,  
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,  
CC allergic diseases, cardiovascular diseases, brain disorders,  
CC degenerative diseases placental, pancreatic, skeletal and muscle  
CC disorders.

Sequence 1289 AA;

Query Match	12.8%	Score 256.5	DB 22	Length 1289
Best Local Similarity	26.8%	Pred. No. 8.2e-10		
Matches 84	Conservative 32	Mismatches 91	Indels 107	Gaps 20

OY	128	CSPG-----	TYBPDLACGGGQRRCGNGHCSDGSHOGGSGRCMHGTV-----	173	
Db	567	cmprytrshpqhrycvdvneeca-----	epcpgyrglcmnty-----gshncryrlylhvqag	640	
OY	174	GPLCTD-----	CMDFYS-----	SLRNETHSICTACDESKTCISGL	209
Db	641	grscvdlneacaphlqdgfcflnfpdyhkencygyrlylkasrpyvedide-----	crdpssc	659	
OY	210	TNRDGE-----	CEVGV- VLDEGCYVDDECAEPPEPCSAOPECCKNANSGYT-----	256	
Db	700	pdgkceakpysfkclacqpgyrrsgsggaacrdvneec- aegspcsgy wcehlpsfrtca		757	
OY	257	-----	EBCDSSCV-----	CCTGEGPCNCK-ECISG--AREHGQCADVDECS	295
Db	758	qgyapapdgrscldvdeceagdyvcdnglcsntlpysfcgcqlsglyshardshsedidecd		817	
OY	296	LAETCVARNKNNYNTPCSYCVCPDPG-----	-----	EETBACVPPAEAE-----	337
Db	818	f-paact--ggdcintngsyrcldcpqhrlvygrkcgqdldecsqdpstclphgackhlgg		874	
OY	338	-----	ATGGSPTQ	346	
Db	875	syvcvcddegflptq	888		

Search completed: September 10, 2002, 11:09:21  
Job time: 463 sec

